Méthodes de codage et d'estimation adaptative appliquées aux communications sans fil

Florence Alberge

Habilitation à diriger des recherches

Méthodes de codage et d'estimation adaptative appliquées aux communications sans fil

Florence Alberge

UNIVERSITÉ PARIS SUD

Table des matières

Exposé détaillé

Publications

Avant-propos

Dans ce manuscrit, je vais décrire mon parcours et les activités de recherche menées depuis mon recrutement à l'Université Paris-Sud. J'évoquerai également, mais de manière moins détaillée, mes travaux antérieurs. Le document est construit en trois parties avec une structure conforme aux usages et recommandations de l'école doctorale STIS de Paris-Sud.

La première partie intitulée *Parcours Professionnel* retrace sous forme de Curriculum Vitae ma formation, mon parcours et décrit l'ensemble des activités et responsabilités exercées. Les activités liées à la recherche y sont évoquées de manière factuelle : liste de publications, collaborations, activités contractuelles ou de valorisation et participation à des co-encadrements de thèse ou de stages de recherche. Le second chapitre est une synthèse de mes activités de recherche dans lequel sont décrits les problématiques, la démarche suivie, les solutions apportées et les choix réalisés. Cette partie se termine par un exposé des perspectives et du projet de recherche.

La partie *Exposé détaillé* met l'accent sur un thème de recherche pour lequel les résultats obtenus sont exposés de manière plus précise et exhaustive que dans la partie précédente. Les travaux présentés portent sur *les techniques itératives de décodage*.

La troisième et dernière partie est un recueil d'une sélection d'articles permettant un aperçu des contributions sur l'ensemble des thèmes de recherche évoqués dans ce manuscrit.

Parcours professionnel et travaux de recherche

Florence Alberge

CV détaillé

Université Paris-Sud
☎ 01 69 33 60 58 (IUT)
☎ 01 69 33 60 81 (IUT - Direction des études)
☎ 01 69 85 17 57 (LSS)
✉ florence.alberge@u-psud.fr
www.l2s.centralesupelec.fr/perso/florence.alberge

Etat Civil

Nom	*Alberge*
Prénom	*Florence*
Date et Lieu de naissance	09 août 1971 à Albi (Tarn)
Statut actuel	Maître de conférences, Université Paris-Sud
Composante	IUT d'Orsay, Département Mesures Physiques
Laboratoire	Laboratoire des signaux et systèmes (LSS), Pôle Télécom et Réseaux. Équipe Théorie de l'information et ses applications
Adresse postale	Laboratoire des Signaux et Systèmes (LSS, UMR CNRS 8506), CentraleSupelec, 3 rue Joliot Curie, 91192 Gif-sur-Yvette

Formation

1996-1999 **Doctorat**, *ENST (Télécom ParisTech)*, Paris, *Mention Très Honorable*.
Spécialité : Signal et Image.
Date de soutenance : 1er décembre 1999.
Sujet : *Développement d'algorithmes adaptatifs utilisables dans un contexte multi-voies.*
Jury : Maurice Bellanger (Président), Madeleine Bonnet (Rapporteur), Dirk Slock (Rapporteur), Pierre Duhamel, Yves Grenier, Philippe Loubaton.

1998 **Agrégation**, *(Candidat libre)*.
Physique option Physique Appliquée.

1995–1996 **DEA**, *Université de Cergy-Pontoise*, *Mention Bien*.
Spécialité : Traitement des Images et du Signal.
Stage à l'Institut National des Télécommunications (Télécom Sud-Paris), Evry.
Tuteur de Stage : J-P. Delmas.
Sujet du stage : *Lois asymptotiques d'estimateurs adaptatifs de sous-espaces propres.*

1993–1996 **Titre d'ingénieur**, *Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA)*, Cergy-Pontoise.
Spécialité : Traitement du Signal

1990–1993 **CPGE**, *Lycée Pierre de Fermat*, Toulouse.

1989–1990 **Baccalauréat**, *Académie de Toulouse*, *Mention Bien*.
Série C.

Parcours

Depuis 2000 **Maître de conférences**, *Univ. Paris-Sud*.

Composante de rattachement : IUT d'Orsay, Département Mesures Physiques.
Laboratoire d'accueil : Laboratoire des Signaux et Systèmes (UMR CNRS 8506).

- Interruption d'activité pour congé maternité : avril à août 2004 puis janvier à mai 2009.
- CRCT : janvier à juin 2012.

1999 – 2000 **ATER**, *Univ. Paris-Sud*.

Composante de rattachement : UFR de Sciences.
Laboratoire d'accueil : Laboratoire des Signaux et Systèmes (UMR CNRS 8506).

Activités d'enseignement

Depuis 2000 **Univ. Paris-Sud (IUT d'Orsay)**, *Service d'enseignement* $\approx$ *240heqTD/an*.

- Liste des enseignements (année 2014-2015)
  - Mathématiques et Traitement du Signal (*Cours,TD,TP*) en DUT 2ème année. Création du cours. Coordinatrice de l'équipe enseignante (S3) : 8 intervenants.
  - Statistiques pour la Métrologie (*Cours,TD,TP*) en DUT 2ème année. Création du cours. Coordinatrice de l'équipe enseignante (S3 et S3 décalé) : 7 intervenants.
  - Traitement d'images (*Cours,TD,TP*) en Lic. Pro ICI. Création du cours.
  - Programmation en Python (*TP*) en DUT 1ère année.
  - Systèmes électriques (*TD*) en DUT 1ère année.
  - Mathématiques (*TD*) en DUT 1ère année.
  - Encadrement de projets, suivi de stages de DUT, suivi d'apprentis en DUT et Lic Pro.
- Polycopiés (année 2014-2015)
  - F. Alberge. *Cours et TD de Mathématiques et Traitement du Signal - Partie 1 : Signaux*, DUT MP 2ème année, 18 pages; *Partie 2 : Analyse de Fourier*, 44 pages; *Partie 3 : Convolution, Echantillonnage et Quantification*, 42 pages.
  - F. Alberge. *Cours, TD et TP de Statistiques pour la métrologie*, DUT MP 2ème année, 52 pages.
  - F. Alberge. *Cours et TD de Traitement d'images*, Lic Pro, 64 pages.
- Autres enseignements (au cours de la période 2000-2014)
  - Probabilité, Statistiques (*Cours,TD*) en DUT 2ème année.
  - Analyse Harmonique et Statistiques (*Cours,TD*) en DUT 2ème année.
  - Informatique d'instrumentation (*Cours,TD,TP*) en DUT 2ème année.
  - Électronique (Conditionnement des Signaux) (*TD,TP*) en DUT 2ème année.
  - Circuits électriques (*TD,TP*) en DUT 1ère année.
  - Programmation en Langage C (*TD,TP*) en DUT 1ère année.
- Projet Pédagogique (*réponse à Appel à Projet pour la Pédagogie, Univ. Paris-Sud*, 2011)
  - Porteurs du projet : F. Alberge et E. Cassan.
  - But du projet : Rénovation salle de TP de Traitement du Signal/Traitement d'images
  - Choix du matériel pédagogique (carte d'acquisition, DSP, module acoustique, ...), rédaction d'appel d'offre (MAPA), suivi des travaux de rénovation.
  - Budget : 90keuros.
- Participation à divers groupes de travail internes au Dpt Mesures Physiques de l'IUT d'Orsay ("Amélioration de la communication : refonte du site web du département", "Semestrialisation", "Suivi des stages en entreprise", "Mise en place des nouveaux programmes : choix des modules complémentaires",...)
- Vulgarisation scientifique (publications)
  - F. Alberge, *"Les courants porteurs en ligne"*, Sciences Ouest, numéro 218, Fev 2005.
  - Co-auteur de l'ouvrage collectif *"Abécédaire de la Physique"*, Centre de Vulgarisation de la Connaissance, CNRS Editions, 2006.

1999 – 2000 **Univ. Paris-Sud (ATER)**, *Volume horaire ≈ 96h*.

- Électronique (*TD,TP*), Polytech Paris-Sud, BAC + 5.
- Traitement du Signal (*TD,TP*), Polytech Paris-Sud, BAC + 5.
- Physique (*TP*), UFR Sciences, L2.
- Projet Professionnel (*TD*), UFR STAPS, L1.

1996 – 1999 **ESME-Sudria et ENST (vacations)**, *Volume horaire ≈ 220h*.

- Traitement du Signal (*Cours, TP*), ESME-Sudria, BAC+4 et BAC+5.
- Filtrage adaptatif et compression des données (*Cours*), ESME-Sudria, BAC +5.
- Filtrage adaptatif (*TD, TP*) , ENST, BAC+4.
- Processeur de Traitement du Signal (*TP*) , ENST, BAC+4.
- Électronique Analogique et Numérique (*TP*) ESME-Sudria, BAC+5.
- Initiation à Matlab (*Cours, TP*), ENST, formation continue.

Activités liées à l'administration

En liaison avec la Recherche

2012-2015 **Membre élu du Conseil National des Universités (CNU)**.

Membre suppléant en 2012 puis titulaire depuis 2013. Participation aux campagnes de qualification MCF, de promotion MCF-HC et de CRCT de 2012 à 2015 inclus. Participation à la campagne PEDR en 2014.

2011 – 2014 **Membre du Conseil de Laboratoire**, *Laboratoire des Signaux et Systèmes*.

Représentante de la Division télécom et réseaux.

2009 – 2012 **Membre élu au Conseil Scientifique (CS)**, *Univ.Paris-Sud*.

Participation au groupe de réflexion "Procédures et fonctionnement du CS", membre de la commission de la pédagogie de l'université Paris-Sud, représentante des rangs B du CS au conseil du département de physique de la faculté d'Orsay (2011-2012). Représentante du CS aux auditions des porteurs de projet BQR (Bonus Qualité Recherche) en math, informatique et sciences humaines et sociales.

Depuis 2011 **Participation à 2 comités de sélection**, *Univ.Paris-Sud*.

MCF-1700, MCF-4121

2007 – 2010 **Membre élu de la commission de spécialiste 61**, *Univ.Paris-Sud*.

- Participation à 4 comités de sélection de MCF (Paris-Sud et Supélec) : MCF-520, MCF-845, MCF-2269, MCF-862
- Participation aux classement des dossiers ATER, demandes de CRCT, professeurs invités...

En liaison avec la Pédagogie

Depuis 2015 **Membre du Conseil Académique**, *Univ. Paris-Saclay*.

Membre élu dans le collège des directeurs et directeurs adjoints des entités de formation.

2012 **Participation au groupe de travail national sur les programmes de Math-Info-TDS**, *DUT Mesures Physiques*.

Ce groupe de travail formé d'une dizaine d'enseignants en DUT Mesures Physiques a eu pour objectif de construire les programmes des trois modules de mathématiques enseignés en 1ère année, des module "Mathématiques et traitement du signal" et "Mathématiques pour la physique" enseignés en deuxième année ainsi que du module "Informatique scientifique". Ces programmes sont appliqués depuis la rentrée 2013.

Depuis 2008 **Directrice des études (DUT $2^{ème}$ année)**, *IUT d'Orsay*, Département Mesures Physiques.

- Responsables : F. Alberge et C. Frappart.
- Contexte : 150 étudiants en moyenne - 2 promotions : S3/S4 et S3 décalé/S4 décalé.
- Rôle : Organisation de l'accueil des étudiants, suivi des notes, suivi des absences, organisation des jurys de validation de semestre, organisation parcours spécifiques à l'étranger, gestion des situations particulières, recrutement des nouveaux entrants en 1ère et 2ème année...

Depuis 2008 **Membre du bureau du département**, *IUT d'Orsay*, Département Mesures Physiques.

2003-2006 **Membre élu du Conseil d'IUT**, *IUT d'Orsay*.

2002 – 2003 **Directrice des études (DUT $1^{ère}$ année)**, *IUT d'Orsay*, Département Mesures Physiques.

- Responsables : F. Alberge et G. Vincents.
- Contexte : 200 étudiants en moyenne - 2 promotions : S1/S2 et S1 décalé/S2 décalé.
- Rôle : Organisation de l'accueil des étudiants, suivi des notes, suivi des absences, organisation des jurys de validation de semestre, gestion des situations particulières, recrutement des nouveaux entrants en 1ère année...

2001 – 2010 **Organisation des projets tutorés**, *IUT d'Orsay*, Département Mesures Physiques.

- Responsables : F. Alberge et C. Frappart.
- Rôle : collecte des sujets, collecte et harmonisation des notes, organisation des jurys de soutenance, présentation du module aux étudiants.

Activités liées à la recherche

Participation à des comités, Editorial boards, organisation de colloques, séminaires...

2015 **Expert pour l'ANR**.

2000 – 2002 **Responsable de l'organisation du séminaire "Télécom"**, *LSS*.

2002 **Membre du comité de programme de la conférence ICC**, *New-York*.
Organisation d'une session technique "Advanced Signal Processing for Communications."

Depuis 2000 **Relectrice d'articles de revue**.
IEEE Trans on Signal Processing, IEEE Signal Processing letters, IEEE Communications letters, IEEE Trans on Communications, EURASIP Journal on Wireless Communications and Networking.

Depuis 2000 **Relectrice d'articles de conférences**.
IEEE International Conference on Communication (ICC), IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP), European Signal Processing Conference (EUSIPCO).

Collaborations académiques

A. Gilloire *CNET, Lannion*

- Thème de recherche : annulation d'écho acoustique pour la téléconférence stéréophonique. Construction d'une procédure de test pour comparer des méthodes d'annulation d'échos stéréophoniques publiées dans la littérature.
- Production : 1 publication dans les actes d'une conférence internationale avec comité de lecture, 1 rapport de recherche.

Y. Grenier *ENST, Paris*

- Thème de recherche : annulation d'écho acoustique pour la téléconférences stéréophonique.
- Production : 2 publications dans les actes de conférence internationales avec comité de lecture, 1 rapport de recherche.

P. Duhamel *ENST, Paris puis CNRS/LSS*

- Collaboration régulière, co-encadrements de thèses et nombreuses publications en commun

M. Nikolova *CMLA, ENS Cachan*

- Thème : Algorithmes adaptatif pour l'égalisation et l'identification de canal.
- Production : 2 publications dans IEEE Trans on Signal Proc. et 3 publications dans les actes de conférences internationales avec comité de lecture.

G. Matz *Univ. de Technologie de Vienne (TU Wien), Autriche*

- Thème 1 : Extension de l'algorithme de Blahut-Arimoto par des techniques issues de la géométrie de l'information.
- Production : 1 rapport de recherche.
- Thème 2 : Décodage itératifs pour les turbo-codes.
- Production : Rédaction de livrables pour le réseau d'excellence Newcom ++. Accueil au LSS d'un doctorant de G. Matz (S. Schwandter, 1 mois), 1 publication conjointe (Z. Naja/S. Schwandter) dans les actes d'une conférence internationale avec comité de lecture.

L. Szczecinski *INRS-EMT, Montreal, Canada*

- Thème : Protocole HARQ pour la sécurité au niveau de la couche physique.
- Production : 1 publication dans les actes d'une conférence internationale avec comité de lecture.

M. Le Treust *ENSEA, Cergy-Pontoise*

- Thème : Protocole HARQ pour la sécurité au niveau de la couche physique.
- Production : 1 publication dans les actes d'une conférence internationale avec comité de lecture, 1 article soumis à *IEEE Trans. on Communications*.

Collaborations industrielles

L. Mazet *Motorola Labs, Gif-sur-Yvette*

- Thème : Estimation et suivi de canal dans un système OFDM. Suite du projet RNRT Festival. Mise en compétition des méthodes développées par Motorola Labs avec celles développées au LSS lors de la thèse de J-M Mamfoubmi.
- Production : 1 rapport de recherche.

A. Rouxel *Wavecom, Issy-les-Moulineaux*

- Thème : Egalisation de canal pour la téléphonie mobile.
- Production : 1 publication dans les actes d'une conférence internationale avec comité de lecture, 1 brevet, 1 rapport de recherche.

A. Ortega-Molina, R. Visoz *Orange Labs, Issy-les-Moulineaux*

- Thème : Allocation de puissance et routage efficace dans les réseaux sans fil. Financement de la thèse de P. Gerold (Contrat de Recherche Externalisé).
- Production : 1 publication dans les actes d'une conférence internationale avec comité de lecture, 1 publication dans les actes d'une conférence nationale avec comité de lecture, 1 article de revue en préparation et 3 rapports de recherche.

Réseaux internationaux et projets

2015 **Nouveau projet soumis : *Robust Integrated Infrastructure for SEcure NETworks (RIISENET)*.**
Réponse à appel à projet CHIST-ERA, FP7, Union européenne.

2012 – 2015 **Participation au Réseau d'excellence européen Newcom♯ en Communications sans fil.**
Participation à deux JRA : *Non-binary codes* et *Towards a backward compatible relaying scheme*.

2008 – 2011 **Participation au Réseau d'excellence européen Newcom++ en Communications sans fil.**
Contribution au délivrable DR4.1 (*Theoretical framework for iterative processing - Geometrical interpretation of iterative algorithms* - Contribution au délivrable DR4.3 (section *Information geometry* et *Bit Interleaved coded modulation with iterative decoding*) en collaboration avec G. Matz (TU, Vienne, Autriche).

Actions de valorisation, brevets

Depuis 2011 **Membre du Comité Stratégique des Brevets (CSB)**, *Univ.Paris-Sud*.
- Rôle : Accompagner l'équipe valorisation de Paris-Sud dans ses décisions sur les suites à donner aux demande de brevets enregistrées (dépôt/extension/abandon).
- Composition : le Vice-président de la Commission de la recherche, les Vice doyens recherche des UFR (droit, pharmacie, médecine, science, STAPS) et cinq représentants des domaines de recherche : M. Mellah (chimie), E. Warde (physique), A. Bennaceur (médecine), Y. Levi (pharmacie) et F. Alberge (sciences de l'ingénieur).

2004 **Brevet**.
R. Gallego, A. Rouxel, F. Alberge, P. Duhamel, *Method for receiving a signal using a maximum likelihood criterion, receiving device and radiotelephone for carrying out said method*, numéro de publication: WO2005096577 (dépôt initial en 2004, PCT en 2005).
Page web : *http://patentscope.wipo.int/search/en/WO2005096577*.
L'invention concerne un procédé de réception d'un signal mis en œuvre dans un système présentant une entrée et au moins deux sorties. Le procédé met en œuvre un critère de type maximum de vraisemblance permettant de minimiser une expression, dite distance contrainte, tenant compte d'une distance entre un vecteur de symboles reçus du signal et un vecteur obtenu par produit matriciel d'un vecteur de symboles émis correspondant et d'une matrice représentative d'un canal de transmission du système.

Encadrement

Co-encadrement de thèse (thèses soutenues)

2011-2014 **Zeina Mheich**, *Durée : 3 ans 9 mois (Octobre 2010 - Juin 2014)*.
- Sujet : *Schémas pratiques pour la diffusion (sécurisée) sur les canaux sans fil*.
- Taux d'encadrement : 50%.
- Publications : 2 publications dans des revues internationales avec comité de lecture (*EURASIP Journal on Wireless Communications and Networking, IEEE Trans. on Communications*), 3 publications dans les actes de conférences internationales avec comité de lecture (*EUSISPCO'13, ICASSP'14, EUSIPCO'14*), 1 article soumis à *IEEE Trans. on Communications*.
- Situation actuelle : en Post-Doc au CEA de Grenoble.

2007-2011 **Z. Naja**, *Durée : 3 ans 6 mois (Octobre 2007 - Avril 2011).*

- Sujet : *Interprétation et amélioration d'une procédure de démodulation itérative.*
- Taux d'encadrement : 50%
- Publications : 1 publication dans une revue internationales avec comité de lecture (*IEEE Trans. on Signal Processing*), 4 publications dans les actes de conférences internationales avec comité de lecture (*ICASSP'09, ICASSP'09, SPAWC'10, ICASSP'11*), 1 publication dans les actes d'une conférence nationale avec comité de lecture (*GRETSI'09*).
- Situation actuelle : Maître de Conférences, Université Libanaise, Tripoli, Liban.

2002-2006 **J-M. Mamfoumbi-Ocloo**, *Durée : 3 ans 9 mois (Octobre 2002 - Juin 2006).*

- Sujet : *Estimation de canaux de transmission respectant un modèle d'évolution temps/fréquence.*
- Taux d'encadrement : 50%.
- Publications : 3 publications dans les actes de conférences internationales avec comité de lecture (*EUSIPCO'04, SPAWC'05, ICASSP'06*).
- Situation actuelle : Directeur de l'ingénierie des réseaux et de la sécurité, Agence Nationale des Infrastructures et des Fréquences (ANINF), Libreville, Gabon.

1999-2002 **F. Abdlekefi**, *Durée : 3 ans (Octobre 1999 - Octobre 2002).*

- Sujet : *Codes de Reed-Solomon pour la correction d'erreurs impulsives dans les systèmes multiporteuses.*
- Taux d'encadrement : 50%.
- Publications : 3 publications dans des revues internationales avec comité de lecture (*IEEE Trans. on Communications)*, 8 publications dans les actes de conférences internationales avec comité de lecture (*ISSPA'01, ICC'02, ICC'03, ISSPA'03, Globecom'03, IST Mobile & Wireless Communications'04, International Symposium on Wireless Personal Multimedia Communications'04, VTC'04*), 2 publications dans les actes de conférences nationales avec comité de lecture (*GRETSI'01, GRETSI'03*).
- Situation actuelle : Maître Assistante, Ecole Supérieure des Télécommunications de Tunis (Sup'Com), Tunisie.

Encadrement au niveau BAC+5

2014 **R. Reyes**, *Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona*, Durée : 6 mois, Taux d'encadrement : 50%.
Sujet : Analyse de la combinaison de codes-espaces-temps et de codes correcteurs d'erreurs (suite du travail de stage de T. Ta).

2014 **V. Brischi Olivatto**, *School of Electrical and Computer Engineering (FEEC), University of Campinas UNICAMP, Brésil*, Durée : 6 mois, Taux d'encadrement : 50%.
Sujet : Techniques de décodage itératif pour les codes LDPC non binaires.

2012 **T. Ta**, *Univ. of Engineering and Technology, Vietnam national university, Hanoi*, Durée : 6 mois, Taux d'encadrement : 50%.
Sujet : Analyse de la combinaison de codes-espaces-temps et de codes correcteurs d'erreurs.

2011 **P. Gerold**, *Master SAR, Supelec - Univ Paris-Sud*, Durée : 6 mois, Taux d'encadrement : 50%.
Sujet : Coopération dans les réseaux sans fil supervisée par les stations de base.

2009 **N. Elpitiya**, *Supelec, Gif-sur-Yvette*, Durée : 3 mois, Taux d'encadrement : 50%.
Sujet : Modélisation du décodage "turbo" comme un problème d'optimisation sous contraintes.

2007-2008 **X. Zhe**, *Electric and Information College North Western Poly-Technology University Xi'an, Chine*, Durée : 1 an, Taux d'encadrement : 50%.
Sujet : Modulations hiérarchiques pour la transmission de vidéo compatible avec le standard DVB-SH.
Publication : 1 publication dans les actes d'une conférence internationale avec comité de lecture (*WCNIS'10*).

2007 **K. Rohan**, *Indian Institute of Technology, Guwahati, Inde*, Durée : 3 mois, Taux d'encadrement : 50%.
Sujet : Analyse de convergence et extensions pour le décodage itératif dans les BICM (Bit Interleaved Coded Modulation).

2002 **K. Amimer**, *DEA traitement du signal et des images, Univ Cergy-Pontoise*, Durée : 6 mois, Taux d'encadrement : 50%.
Sujet : Interprétation d'une procédure de démodulation itérative en termes de projection et de minimisation de distance dans des espaces de densité de probabilité.

Déroulement de carrière

Mon travail en tant que chercheur a débuté en 1996 lors de mon stage de DEA à l'INT sous la direction de Jean-Pierre Delmas. J'ai alors étudié des algorithmes d'analyse en composante principale introduits dans la littérature des réseaux neuronaux avec des résultats asymptotiques et en terme de vitesse de convergence.
J'ai ensuite poursuivi par une thèse à l'ENST, soutenue en décembre 1999, pendant laquelle j'ai travaillé sur des algorithmes adaptatifs en présence de signaux corrélés avec application au domaine des télécommunications. Cette thèse a été dirigée par Pierre Duhamel et co-encadrée par Yves Grenier. Souhaitant enseigner, j'ai fait des vacations à l'ESME-Sudria ainsi qu'à l'ENST pour un volume horaire légèrement supérieur à celui d'un monitorat. En parallèle, en 98, j'ai fait le choix de m'inscrire à l'agrégation de physique, option physique appliquée, en tant que candidate libre. J'ai obtenu l'agrégation cette même année.
J'ai ensuite obtenu un (demi) poste d'ATER à l'Université Paris-Sud en 1999 avec un rattachement à l'UFR de Sciences pour l'enseignement et au Laboratoire des Signaux et Systèmes (LSS) pour la recherche. Suite à un départ et à la retraite et à une démission, l'équipe Télécommunications du LSS se trouvait réduite à cette époque là à deux personnes. J'ai été recrutée en tant que maître de conférences l'année suivante à l'Université Paris-Sud avec l'IUT d'Orsay comme composante de rattachement et le LSS comme laboratoire de recherche avec pour objectif de contribuer au maintien d'une activité en Télécom au LSS. La même année Pierre Duhamel a obtenu un poste de directeur de recherche au LSS. J'ai donc logiquement, au cours des années qui suivent, continué de collaborer avec Pierre Duhamel principalement par le biais de co-encadrements de thèses.

Principaux cours enseignés et pédagogie

Mon service d'enseignement fluctue, selon les années, entre 220h et 270h équivalent TD avec une moyenne autour de 240h équivalent TD. A l'heure actuelle, les matières que j'enseigne sont le traitement du signal, le traitement d'images et les statistiques. J'ai la charge du cours pour ces trois thèmes et j'anime donc aussi l'équipe pédagogique. Je participe également en tant que chargée de TD et/ou chargée de TP aux enseignements de mathématiques, systèmes électriques et informatique. J'interviens dans les trois année de formation : DUT 1ère et 2ème année et licence professionnelle.
Le module "Traitement d'images" en licence professionnelle est un enseignement que j'ai proposé et entièrement crée et qui est dispensé aux étudiants depuis l'ouverture de la licence en 2005. Il a subi différentes évolutions depuis sa création pour s'adapter au public et aussi aux évolutions de la licence elle-même.
Le cours de "Mathématiques et Traitement du Signal" fait partie du tronc commun de la 2ème année du DUT Mesures Physiques. Le programme est national et commun à tous les DUT Mesures Physiques. J'ai participé au groupe de travail "Math-Info-TDS" chargé de construire les nouveaux programmes suite à la dernière réforme du baccalauréat. Ces programmes sont appliqués depuis la rentrée 2013. J'ai introduit, dans le cours que je donne à l'IUT d'Orsay, de nombreux exemples issus de la physique (optique, acoustique,

électronique,...) afin de relier cet enseignement aux autres disciplines et afin d'illustrer par des applications des concepts parfois difficiles à appréhender. Je présente en général les notions abordées de manière symbolique (équations), schématique, visuelle et/ou sonore dans une volonté de m'adapter à un public hétérogène. En TP, j'ai souvent recours aux signaux audio et/ou aux images ce qui permet de travailler, de comprendre et de réfléchir sur un support concret et ludique. Cette approche donne de bons résultats y compris auprès des étudiants les moins motivés et les plus faibles en mathématiques.
Le cours de "Statistiques pour la métrologie" fait partie d'un module global intitulé "Métrologie, Qualité, Statistiques". J'ai en charge la partie "Statistiques" alors que la partie "Métrologie, Qualité" est enseignée par un professeur associé (PAST) travaillant dans l'industrie. Nous avons travaillé conjointement afin de construire un contenu cohérent. J'ai choisi évidemment dans mon cours des exemples centrés sur les problématiques de la mesure.
En 2011, Eric Cassan et moi-même avons été porteurs d'un projet de rénovation pédagogique. Ce projet a été sélectionné par l'université Paris-Sud et a obtenu un financement de 90keuros. Nous avons pu complètement rénover une salle de TP que nous utilisons en traitement du signal et des images : achat de mobilier permettant une utilisation de la salle en configuration TD ou TP, achat de nouveaux ordinateurs, câblage électrique et réseau et achat de matériel pédagogique (cartes d'acquisition, DSP, modules pour l'acoustique, casques, caméras, ...).
Depuis 2008, j'assure avec mon collègue Claude Frappart la direction des études de la deuxième année du DUT. Nous nous occupons de la deuxième année classique (100 étudiants environ, septembre de l'année $n-1$ à juillet de l'année n) et de la deuxième année décalée (50 étudiants environ, février de l'année $n-1$ à janvier de l'année n). Nous consacrons en moyenne une demi-journée par semaine à cette responsabilité. Notre travail se divise entre suivi des étudiants (absence, niveau, comportement,...), gestion administrative (jurys, notes, admissions,...) et organisation des parcours pédagogiques (MCC, options, ...).
La pédagogie m'a toujours intéressée, j'ai participé au cours de ma carrière à un grand nombre de groupes de réflexions sur notre offre de formation et/ou notre pratique pédagogique. Nous avons la chance à l'IUT d'Orsay d'avoir une "cellule de réflexion pédagogique" animée par plusieurs collègues dynamiques et enthousiastes qui proposent des ateliers ou des cycles de conférence nous permettant de mettre en commun nos pratiques pédagogiques, de les remettre en question et de nous tenir informés des dernières avancées en didactique.

Synthèses des travaux, résultats principaux et projet de recherche

Je présente ci-après les thématiques de recherche auxquelles je me suis intéressée par ordre chronologique. Elles seront plus amplement développées dans le chapitre 2.

○ Algorithmes adaptatifs et application (Stage DEA, Thèse (1996-1999))

Le premier volet de ce travail a été réalisé lors de mon stage DEA à l'INT avec pour objectif l'étude d'algorithmes d'analyse en composante principale introduits dans la littérature des réseaux neuronaux. Nous avons montré que, la plupart de ces algorithmes ont des similarités avec les algorithmes stochastiques de type gradient rencontrés en traitement du signal. Nous avons établi les lois asymptotiques de différents estimateurs de structures propres et nous avons obtenu des expressions analytiques des matrices de covariance des vecteurs propres estimés et/ou des matrices de projection associées très similaires à celles rencontrées dans les algorithmes bloc. Enfin, les vitesses de convergence et les déviations par rapport à l'orthogonalité ont été comparées permettant de déterminer le meilleur compromis.
J'ai également travaillé sur les algorithmes adaptatifs lors de ma thèse à l'ENST mais dans un contexte applicatif différent. Nous nous sommes intéressés à l'annulation d'écho stéréophonique et à l'égalisation aveugle dans un système mono-entrée/multi-sorties. Dans les deux cas, les signaux reçus sur les capteurs (micro, antenne) sont des combinaisons linéaires d'une même source et sont donc fortement corrélés. En annulation d'écho, la réponse impulsionnelle du canal à identifier comporte un nombre important d'échantillons significatifs imposant un algorithme d'estimation à faible complexité. Nous avons montré comment l'utilisation conjointe d'un algorithme du gradient, de bancs de filtres et l'introduction d'un pas d'adaptation variable par sous-bande conduit à une solution satisfaisante pour cette application. Nous avons ensuite abordé l'égalisation aveugle par le biais de méthodes d'estimation conjointe du canal et des

symboles. Nous avons choisi une technique de type Maximum de Vraisemblance Déterministe. Nous avons proposé un algorithme adaptatif qui possède les caractéristiques suivantes : faible complexité, convergence rapide, capacité de poursuite des variations du canal et robustesse à la surestimation de l'ordre du canal.

Bilan : 2 publications dans des revues internationales avec comité de lecture ([R.9],[R.10]), 5 publications dans les actes de conférences internationales avec comité de lecture ([C.25] à [C.29]), 1 publication dans les actes d'une conférence nationale avec comité de lecture [N.6], 1 rapport de recherche [Ra.4].

- Codage correcteur de bruit impulsionnel (2000-2004)

Nous nous plaçons dans le contexte des systèmes multi-porteuses de type OFDM (Orthogonal Frequency Division Multiplexing). L'objectif est ici de trouver une méthode efficace pour lutter contre les perturbations de type bruit impulsionnel dues à des commutations de courant par exemple. La méthode de correction que nous avons exploitée repose sur une analogie entre le système OFDM le plus simple (à base de transformée de Fourier rapide) et les codes Reed-Solomon dans le corps des complexes. Les codes Reed-Solomon sont des codes cycliques correcteurs d'erreurs multiples. L'analogie repose sur l'observation suivante. Pour un modulateur numérique, on calcule souvent une version sur-échantillonnée du signal à temps continu ce qui revient à ajouter des zéros à la séquence de symboles en entrée du modulateur. Ces zéros sont l'équivalent des syndromes du code et permettent de contrôler l'apparition éventuelle d'erreurs lors de la transmission. Cette façon de poser le problème est décrite dans les travaux de Wolf en 1983 ainsi que dans ceux de Redinbo en 2000. Ces contributions considèrent que les séquences émises contiennent un nombre suffisant de zéros consécutifs. Cette condition n'est pas pertinente ici puisque les zéros en question sont situés sur une partie du spectre qui est atténuée par les filtres de mise en forme. Une des originalités de notre travail est d'avoir montré que l'utilisation de symboles pilotes (à la place des zéros) en tant que syndromes du code est également possible. Notre approche s'applique aux situations où les symboles connus ne sont ni consécutifs ni régulièrement répartis dans la séquence ce qui généralise les résultat existants dans la littérature. Nous avons ensuite montré que la capacité de correction dépend de la position des syndromes au sein de la séquence. Nous avons établi une condition nécessaire pour une capacité de correction maximale qui étend les bornes déjà connues.
Nous avons ensuite considéré les problèmes d'écrêtement (clipping) du signal OFDM. C'est un problème fréquemment rencontré puisque le signal OFDM a la particularité d'avoir une dynamique importante. Nous avons montré que ce problème peut être vu comme un problème de correction d'erreurs impulsionnelles et que notre méthode permet de réduire le niveau de PAPR (Peak to Average Power Ratio). Les recherches dans ce domaine se sont intensifiées au cours des dernières années; nos travaux sont bien référencés dans la littérature récente. L'article [R.8] par exemple est cité 80 fois dont la moitié dans les 5 dernières années.

Bilan : 1 thèse soutenue (F. Abdelkefi), 3 publications dans des revues internationales avec comité de lecture ([R5], [R6], [R8]), 8 publications dans les actes de conférences internationales avec comité de lecture ([C16] à [C18] et [C20] à [C24]) et 2 publications dans les actes de conférences nationales avec comité de lecture ([N.4] et [N.5]).

- Identification aveugle ou semi-aveugle de canaux de communication (2002-2008)

Une partie de ce travail est la suite des travaux en égalisation aveugle commencés lors de la thèse. La nouveauté par rapport aux travaux antérieurs est l'introduction d'un *a priori* sur les symboles dans le critère à optimiser, l'étude de l'influence de l'*a priori* sur la présence de minima locaux et le passage à un algorithme adaptatif à faible coût et à convergence rapide. Dans le cadre d'une collaboration avec la société Wavecom, nous avons adapté la technique précédente à une problématique d'égalisation semi-aveugle pour un système de téléphonie mobile de type GSM pour des situations à haute vitesse. Pour cela nous avons introduit un critère reposant sur le partitionnement du paquet de données. La matrice de canal est alors mieux conditionnée ce qui améliore les performances tout en diminuant la complexité calculatoire de la méthode d'estimation. La version adaptative proposée a des performances comparables aux méthodes de

référence pour la faible à moyenne vitesse et des performances supérieures à haute vitesse. Elle s'applique à tout système utilisant une modulation GMSK (Gaussian minimum-shift keying).
La deuxième partie se situe dans le cadre des systèmes multi-porteuses pour des applications potentiellement mobiles. On considère ici encore un solution semi-aveugle. La différence avec les situations décrites plus haut est la nature du canal qui est ici doublement sélectif, en temps et en fréquence, avec un modèle d'évolution partiellement connu. D'un point de vue méthodologique, nous avons choisi de travailler avec l'algorithme Expectation-Maximization (EM) qui a l'avantage d'avoir une complexité algorithmique linéaire. Pour prendre en compte l'*a priori* sur le canal (le modèle de corrélation), nous remplaçons la maximisation au sens du maximum de vraisemblance par une maximisation *a posteriori* (EM-MAP); l'effet bénéfique sur la qualité de l'estimateur est contrebalancé par l'augmentation de la complexité arithmétique qui devient quadratique. Nous avons proposé une approximation de l'algorithme EM-MAP par une méthode de type "One Step Later" (Green, 1990). L'algorithme résultant a pour propriété : une complexité calculatoire linéaire, une convergence monotone et garantie vers un point stationnaire de la probabilité *a posteriori*.

Bilan : 1 thèse soutenue (J-M Mamfoumbi), 1 publication dans une revue internationale avec comité de lecture [R.7], 1 brevet [B.1], 5 publications dans les actes de conférences internationales avec comité de lecture ([C.12] à [C.15] et [C.19]) et 2 rapports de recherche ([Ra.2] et [Ra.3]).

- Techniques itératives de décodage pour les turbo-codes et les LDPC (2007-)

Le décodage itératif consiste à décoder un code complexe à l'aide de plusieurs décodeurs plus simples qui échangent de l'information (extrinsèques) au fil des itérations. L'apparition des turbo-codes en 1993 a permis de montrer que des performances proches de la capacité pouvaient être atteintes en procédant à un décodage itératif au niveau du récepteur. Bien que relativement simple à mettre en oeuvre, le décodage itératif est extrêmement difficile à analyser puisqu'il n'a pas été introduit à l'origine comme la solution d'un problème d'optimisation. C'est pourtant sous l'angle de l'optimisation que nous avons étudié le décodage itératif. Nous avons établi le mécanisme permettant d'obtenir les équations classiques du décodage itératif à partir du maximum de vraisemblance sur la séquence complète de taille n. Il procède en deux étapes : passage d'un critère global à une suite de n critères locaux puis implémentation de type Jacobi/Gauss-Seidel. Le propagation d'extrinsèques est une conséquence de cette implémentation. En utilisant la théorie des jeux, nous avons proposé une interprétation nouvelle dans laquelle la moyenne des n critères locaux est la fonction d'utilité et où un point fixe de l'algorithme est un équilibre de Nash. Nous avons prouvé que la convergence vers un équilibre de Nash pouvait toujours être obtenue en adaptant la quantité d'information (extrinsèque) échangée entre les décodeurs.
Les performances des turbo-codes et des codes LDPC sont souvent analysées à l'aide d'EXIT charts (EXtrinsic Information Transfer) qui sont une représentation graphique de l'évolution de l'information (niveau de connaissance) acquise par chaque décodeur au cours du processus itératif. Le tracé d'un EXIT chart nécessite la connaissance du message transmis et est utilisé pour prédire la performance du système mais ne peut pas être utilisé au niveau du récepteur. Dans ce contexte, nous nous sommes intéressés à l'information mutuelle calculée à partir des extrinsèques (et non plus du message). Nous montrons le lien entre cette quantité et celle tracée dans les EXIT charts. Nous montrons ensuite comment cette quantité peut être évaluée au niveau du récepteur. Nous montrons enfin comment utiliser cette mesure pour un auto-diagnostic au niveau du récepteur en terme de critère d'arrêt, de détection d'erreurs et d'adaptation de la quantité d'information à échanger entre les décodeurs.

Bilan : 1 thèse soutenue (Z. Naja), deux publication dans des revues internationales avec comité de lecture ([R.1] et [R.3]), 6 publications dans les actes de conférences internationales avec comité de lecture ([C.4], [C.6], [C.7] et [C.9] à [C.11]), 2 publication dans les actes d'une conférence nationale avec comité de lecture ([N.1] et [N.3]).

◦ Diffusion sur les canaux sans fil (2011-)

Ce travail se situe dans le domaine des communications multi-utilisateurs. Nous nous intéressons aux systèmes de diffusion (un émetteur / plusieurs récepteurs) sur les canaux sans fil. Les limites théoriques de communication sont connues mais sont atteignables par des schémas et structures de communication non-implémentables en pratique. Dans les systèmes implémentés, l'alphabet est de taille finie et les symboles transmis sont en général équiprobables. L'objectif de cette étude est l'évaluation de l'impact des contraintes d'implémentation sur les débits atteignables. Nous avons considéré deux types de canaux de diffusion gaussiens à deux utilisateurs. Dans le premier cas, l'émetteur envoie un message commun aux deux utilisateurs et un message privé vers l'un des deux utilisateurs alors que dans le second cas le message privé est aussi un message confidentiel pour lequel la sécurité doit être garantie. Dans chacun des cas, nous avons étudié le compromis entre complexité d'implémentation et efficacité et sommes en mesure de déterminer la stratégie à utiliser pour une situation/application donnée. Nous avons également évalué l'impact de la contrainte de sécurité en comparant les débits atteignables obtenus dans les deux cas traités. Enfin, nous avons étudié les communications sécurisées avec schéma d'accusé de réception hybride (HARQ) pour un canal à écoute à évanouissement par bloc. Nous avons supposé une connaissance imparfaite du canal à l'émetteur reposant uniquement sur les statistiques accompagnée d'une connaissance parfaite des états passés du canal vers le décodeur légitime obtenue à l'aide d'un canal de retour. Dans ce contexte, nous avons mis au point un algorithme permettant de déterminer la politique optimale d'adaptation du débit. Nous montrons que le schéma proposé a un gain significatif en terme de débit utile par rapport aux schémas non adaptatifs proposés dans la littérature.

Bilan : 1 thèse soutenue (Z. Mheich), 2 publications dans des revues internationales avec comité de lecture ([R.2], [R.4]), 3 publications dans les actes de conférences internationales avec comité de lecture ([C.1] à [C.3]), 1 article soumis à IEEE Trans. on Communications [S.1].

◦ Projet de recherche

Les thématiques que je souhaite développer dans les années futures sont dans le domaine du codage, de la sécurité et de la coopération dans les réseaux sans fil mais également du domaine du traitement du signal au sens large.
Le premier axe de recherche a été initié avec la thèse de H. N. Nguyen débutée en avril 2014. Nous nous intéressons à des stratégies de coopération de type relayage dans laquelle l'intégrité des signaux reçus est protégée à l'aide d'un code détecteur d'erreur. Les travaux de Vu en 2012 ont montré l'intérêt du codage canal en terme d'efficacité spectrale et ont prouvé l'utilité du relayage indépendamment de la position des relais. L'objectif de cette thèse est de prolonger et généraliser ce type de résultat en considérant des techniques itératives de passage de message sur les codes concaténés. Nous nous intéresserons également à la problématique de la sélection de relais. Enfin, nous travaillerons sur des situations où les émetteurs n'ont pas connaissance de l'existence de relais et où le récepteur adapte ses algorithmes en fonction de la situation choisie (compatibilité ascendante). Cette situation peut être considérée comme modèle d'une voie montante vers une station de base.
Le deuxième axe est liée à la sécurité. Les résultats obtenus dans la thèse de Z. Mheich montrent comment, pour une situation de diffusion d'information avec contrainte de sécurité, le débit de la source peut être adapté (ici en fonction de l'état passé du canal entre émetteur et utilisateur légitime) afin de garantir la sécurité et l'intelligibilité du message pour le décodeur légitime. Ce problème a été abordé sous l'angle de la théorie de l'information. Il sera abordé dans ce projet sous un angle pratique. Les questions traitées seront la construction de codes pour la sécurité, l'adaptation du débit de transmission en prenant en compte les spécificités du canal et les erreurs commises dans l'estimation des paramètres du canal et/ou dans les canaux de retour. La question de la compatibilité avec les systèmes sans fil déjà déployés est un point crucial qui devra aussi être étudié. L'introduction dans LTE du relayage et des services ou communications mobile-à-mobile sont autant de situations qui nécessitent une prise en compte et une étude spécifique en terme de sécurité.

Le troisième axe de recherche est du domaine de l'optimisation pour les problèmes de grande dimension. C'est une thématique en plein essor en traitement du signal. De nombreux algorithmes ont été développés ou remis au goût du jour dans ce cadre comme par exemple les algorithmes adaptatifs, distribués, les méthodes stochastiques pour résoudre des problèmes convexes et non-convexes. Les propriétés que doivent partager ces solutions algorithmiques pour les problèmes de grande dimension sont une complexité calculatoire très faible et la possibilité d'un calcul parallèle ou distribué. Même si certains algorithmes classiques semblent bien adaptés à cette situation, les solutions vraiment efficaces seront probablement mixtes et devront s'adapter à des fonctions de coût ou d'objectif complexes, à des données bruitées et à des contraintes imposées par la structure du problème. Ce sont ces directions de recherche qui seront explorées en priorité.

Liste de publications

L'exposant ^{D} indique un doctorant, ^{S} indique un étudiant en stage de master et ^{P} indique un post-doctorant.

Publications dans des revues internationales avec comité de lecture

[R.1] F. Alberge, On some Properties of the Mutual Information between Extrinsics with Application to Iterative Decoding - *IEEE Transactions on Communications*, vol 63, no 5, p. 195-205, 2015.

[R.2] Z. Mheich^{D}, F. Alberge, P. Duhamel, Achievable Secrecy Rates for the Broadcast Channel with Confidential Message and Finite Constellation Inputs - *IEEE Transactions on Communications,* vol 63, no 1, p. 195-205, 2015.

[R.3] F. Alberge, Z. Naja^{D}, P. Duhamel, Power extrinsic propagation for turbo-codes - *IEEE Transactions on Signal Processing,* vol. 61, no 5, p.1107-1111, 2013.

[R.4] Z. Mheich^{D}, F. Alberge, P. Duhamel, Achievable Rates Optimization For Broadcast Channels Using Finite Size Constellations Under Transmission Constraints - *EURASIP Journal on Wireless Communications and Networking*, p.1-15, 2013.

[R.5] F. Abdelkefi^{D}, P. Duhamel, F. Alberge and J. Ayadi. On the use of cascade structure to correct impulsive noise in multicarrier systems. *IEEE Transactions on Communications,* vol. 56, no. 11, pp. 1844-1858, 2008.

[R.6] F. Abdelkefi^{D}, P. Duhamel and F. Alberge. A necessary condition on the location of pilot tones for maximizing the correction capacity in OFDM systems. *IEEE Transactions on Communications,* vol. 55, no. 2, pp. 356-366, 2007.

[R.7] F. Alberge, M. Nikolova and P. Duhamel. Blind Identification/Equalization using Deterministic Maximum Likelihood and a partial prior on the input. *IEEE Transactions on Signal Processing,* vol. 54, no. 2, pp. 724-737, 2006.

[R.8] F. Abdelkefi^{D}, P. Duhamel and F. Alberge. Impulsive noise cancellation in multicarrier transmission. *IEEE Transactions on Communications,* vol. 53, no. 1, pp. 94-106, 2005.

[R.9] F. Alberge, P. Duhamel and M. Nikolova. Adaptive Solution For Blind Identification/Equalization Using Deterministic Maximum Likelihood. *IEEE Transactions on Signal Processing,* vol. 50, no. 4, pp. 923-936, 2002.

[R.10] J-P. Delmas and F. Alberge. Asymptotic performance analysis of subspace adaptive algorithms introduced in the neural network literature. *IEEE Transactions on Signal Processing,* vol. 46, no. 1, pp. 170-192, 1998.

Brevet

[B.1] R. Gallego^{D}, A. Rouxel, F. Alberge and P. Duhamel. *Procédé de réception d'un signal mettant en oeuvre un critère de type maximum de vraisemblance. Dispositif de réception et radiotéléphone correspondants*. Déposé le 24/03/2005. Identifiant PCT/FR05/00710.

Conférences internationales avec comité de lecture et actes

[C.1] Z. Mheich^{D}, M. Le Treust^{P}, F. Alberge, P. Duhamel, L. Szczecinski. Rate-adaptive secure HARQ protocols for block-fading channels. *EUropean SIgnal Processing COnference (EUSIPCO)*, Lisbon, Portugal, Sept. 2014.

[C.2] Z. Mheich[D], F. Alberge, P. Duhamel. The impact of finite-alphabet input on the secrecy-achievable rates for broadcast channel with confidential message. *Proc. of the IEEE International Conference on Audio, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014. Taux d'acceptation : 48%.

[C.3] Z. Mheich[D], F. Alberge, P. Duhamel. On the Efficiency of Transmission Strategies for Broadcast Channels Using Finite Size Constellations. *EUropean SIgnal Processing COnference (EUSIPCO)*, Marrakech, Morocco, Sept. 2013.

[C.4] F. Alberge. Pairwise joint probability propagation in BICM-ID. *Proceedings of the 7th International Symposium on Turbo Codes and Iterative Information Processing*, Goteborg, Sweden, August 2012.

[C.5] P. Gerold[D], T. Pham[D], F. Alberge, P. Duhamel. Multi-hop, multi-route power minimisation in ad hoc network. *Proc. of the IEEE International Conference on Audio, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012. Taux d'acceptation : 49%.

[C.6] F. Alberge, A game-theoretic interpretation of iterative decoding. *European Signal Processing Conference (EUSIPCO)*. Barcelona, Spain, August 2011. Taux d'acceptation : 52%.

[C.7] F. Alberge, Z. Naja[D] and P. Duhamel. From Maximum Likelihood to iterative decoding. *Proc. of the IEEE International Conference on Audio, Speech, and Signal Processing (ICASSP)*. Prague, Czech Republic, May 2011. Taux d'acceptation : 49%.

[C.8] X. Zhe[S], W. YongSheng, F. Alberge and P. Duhamel. A turbo iteration in 16QAM hierarchical modulation. In *Proc. of the IEEE International Conference on Wireless Communications Networking and Information Security (WCNIS)*. Beijing, China, June 2010.

[C.9] F. Alberge, Z. Naja[D] and P. Duhamel. New criteria for iterative decoding. In *Proc. of the IEEE International Conference on Audio, Speech, and Signal Processing (ICASSP)*. Taipei, Taiwan, April 2009. Taux d'acceptation : 44,7%.

[C.10] Z. Naja[D], F. Alberge and P. Duhamel. Geometrical interpretation and improvements of the Blahut-Arimoto's algorithm. In *Proc. of the IEEE International Conference on Audio, Speech, and Signal Processing (ICASSP)* Taipei, Taiwan, April 2009. Taux d'acceptation : 44,7%.

[C.11] F. Alberge. Iterative decoding as Dykstra's algorithm with alternate I-projection and reverse I-projection. In *EUropean SIgnal Processing COnference (EUSIPCO)* Lausanne, Switzerland, August 2008. Taux d'acceptation : 50%.

[C.12] F. Alberge. Accelerated linear EM-MAP algorithm for OFDM channel estimation. In *Proc. of International Conference on Acoustic Speech and Signal Processing*. Honolulu, USA, April 2007. Taux d'acceptation : 46,25%.

[C.13] J-M. Mamfoumbi Ocloo[D] and F. Alberge. OFDM channel estimation by a linear EM-MAP algorithm. In *Proc. of International Conference on Acoustic Speech and Signal Processing*. Toulouse, France, May 2006.

[C.14] J-M. Mamfoumbi Ocloo[D], F. Alberge and P. Duhamel. Semi-Blind channel estimation for OFDM systems via an EM-MAP algorithm. In *Proc. of Signal Processing Advances in Wireless Communication (SPAWC)*. New York, USA, June 2005.

[C.15] S. Touati[P], J-M. Mamfoumbi Ocloo[D], F. Alberge and P. Duhamel. Semi-Blind channel estimation for OFDM systems via an EM-Block algorithm. In *Proc. of EUropean SIgnal Processing COnference (EUSIPCO)* Vienna, Austria, Sept. 2004.

[C.16] F. Abdelkefi[D], P. Duhamel, F. Alberge and J. Ayadi. Using Pilot Tones Distribution For Maximal Correction Capacity Of Impulse Noise in OFDM Systems. In *Proc. of IEEE Vehicular Technology Conference Fall* Los Angeles, USA, Sept. 2004.

[C.17] F. Abdelkefi[D], P. Duhamel, F. Alberge and J. Ayadi. Correction Capacity of Impulse Noise in Multicarrier Systems. In *Proc. of 7th International Symposium on Wireless Personal Multimedia Communications*. Abano Terme, Italy, Sept. 2004.

[C.18] F. Abdelkefi[D], P. Duhamel, F. Alberge and J. Ayadi. An Efficient Algorithm for the PAPR Level Reduction in OFDM Based Systems. In *Proc. of IST Mobile & Wireless Communications*. Lyon, France, Sept. 2004.

[C.19] R. Gallego[D], F. Alberge, P. Duhamel and A. Rouxel. Semi-blind equalization for GMSK-based mobile communication. In *Proc. of International Conference on Acoustic Speech and Signal Processing (ICASSP)*. Montreal, Quebec, May 2004.

[C.20] F. Abdelkefi[D], P. Duhamel, and F. Alberge. A necessary condition on the location of pilot tones for impulse noise cancellation in OFDM system and its applications in Hiperlan 2. In *Proc of Globecom*. San Francisco, USA, Dec. 2003. Taux d'acceptation : 36,3%.

[C.21] F. Abdelkefi[D], P. Duhamel, and F. Alberge. Improvement of the complex Reed Solomon decoding with application to impulse noise cancellation in Hiperlan 2. In *Proc. of ISSPA*. Paris, France, Aug. 2003.

[C.22] F. Abdelkefi[D], P. Duhamel, and F. Alberge. Impulse noise correction in Hiperlan 2: improvement of the decoding algorithm and application to PAPR reduction. In *Proc. of International Conference on Communication (ICC)*. Anchorage, Alaska, May 2003. Taux d'acceptation : 37,5%.

[C.23] F. Abdelkefi[D], P. Duhamel, and F. Alberge. On the location of pilot tones for impulse noise cancellation in multicarrier transmission. In *Proc. of International Conference on Communication (ICC)*. New York, USA, May 2002. Taux d'acceptation : 41,8%.

[C.24] F. Abdelkefi[D], P. Duhamel, and F. Alberge. On the use of pilot tones for impulse noise cancellation in Hiperlan 2. In *Proc. of ISSPA*. Kuala Lumpur, Malaysia, Aug. 2001.

[C.25] F. Alberge, P. Duhamel and M. Nikolova. Low cost adaptive algorithm for blind channel identification and symbol estimation. In *Proc of EUropean SIgnal Processing COnference (EUSIPCO)*. Tampere, Finland, Sept. 2000.

[C.26] F. Alberge, M. Nikolova and P. Duhamel. Adaptive deterministic maximum likelihood using a quasi-discrete prior. In *Proc of International Conference on Acoustic Speech and Signal Processing (ICASSP)*. Istanbul, Turkey, May 2000.

[C.27] F. Alberge, P. Duhamel and M. Nikolova. Blind identification/equalization using deterministic maximum likelihood and a partial information on the input. In *Proc of Signal Processing Advances in Wireless Communications (SPAWC),* Annapolis, USA, May 1999.

[C.28] F. Alberge, P. Duhamel and Y. Grenier. Comparison of adaptive algorithms for stereophonic acoustic echo cancellation. In *Proc of the* 8^{th} *IEEE Digital Signal Processing Workshop*. Bryce Canyon, USA, Aug. 1998.

[C.29] F. Alberge, P. Duhamel and Y. Grenier. A combined FDAF/WSAF algorithm for stereophonic acoustic stereo cancellation. In *Proc. of International Conference on Acoustic Speech and Signal Processing (ICASSP)*. Seattle, USA, 1998.

- Conférences nationales avec comité de lecture et actes

[N.1] F. Alberge. Propriétés et applications de l'information mutuelle entre extrinsèques. *Actes du GRETSI*, Lyon, France, Sept. 2015.

[N.2] P. Gerold[D], P. Duhamel, F. Alberge, Allocation de puissance en vue d'un routage efficace dans les réseaux sans fil. *Actes du GRETSI*, Brest, France, Sept. 2013.

[N.3] Z. Naja[D], F. Alberge and P. Duhamel. Méthode du point proximal: principe et applications aux algorithmes itératifs. *Actes du GRETSI*. Dijon, France, Sept. 2009. Taux d'acceptation 75%.

[N.4] F. Abdelkefi[D], P. Duhamel, and F. Alberge. Tests en cascade pour la correction des erreurs impulsives et la réduction du PAPR dans le contexte d'Hiperlan 2. *Actes du GRETSI*. Paris, France, Sept. 2003.

[N.5] F. Abdelkefi[D], P. Duhamel, and F. Alberge. Codage correcteur d'erreur dans le corps des complexes et systèmes multiporteuses. *Actes du GRETSI*. Toulouse, France, Sept. 2001.

[N.6] J-P. Delmas and F. Alberge. Lois asymptotiques d'estimateurs adaptatifs de sous-espaces introduits dans la littérature neuronale. *Actes du GRETSI*. Grenoble, France, Sept., 1997.

- Rapports de recherche

[Ra.1] G. Matz, F. Alberge, Z. Naja[D], P. Duhamel. Acceleration the Arimoto-Blahut algorithm via information geometry. 2011.

[Ra.2] F. Alberge, J-M. Mamfoumbi[D], P. Duhamel. Semi-blind estimation of slowly varying channels for OFDM systems via EM-based algorithms. 2007.

[Ra.3] R. Gallego[D], F. Alberge, P. Duhamel, A. Rouxel. Semi-blind equalization for high-speed GMSK-based mobile communications. 2004.

[Ra.4] F. Alberge, P. Duhamel, Y. Grenier. Comparaison des performances d'algorithmes adaptatifs dans un contexte d'annulation d'écho stéréophonique. 1998.

- Autres communications

[Co.1] T.X. Vu, M. El Soussi, H.N. NGuyen[D], P. Duhamel, F. Alberge, L. Vandendorpe. Semi-orthogonal MARC with half-duplex relaying: A backward compatible cooperative network. *JNCW15, Final Newcom♯ event*, Barcelona, Spain, 14-15 Oct. 2015.

[Co.2] P. Duhamel, F. Alberge, H. N. Nguyen[D], M. El Soussi, L. Vandendorpe. (Network coding) for the relay channel : Parial Realying and Backward Compatible Scheme. *Newcom♯, Dissemination Event, Ericsson*, 4th May 2015.

[Co.3] Z. Naja[D], F. Alberge, P.Duhamel, L. Szczecinski. Bit-Interleaved Coded Modulation with Iterative Decoding/Demapping (BICM-ID). *Tutorial in the NEWCOM++ 2010 Winter School on Iterative Techniques in Wireless Communications*, Department of Communication Technology, Aalborg University, Aalborg, Denmark, 24-26 February 2010.

- Travaux récents soumis

[S.1] Z. Mheich[D], M. Le Treust, F. Alberge, P. Duhamel. Rate adaptation for incremental redundancy secure HARQ. Soumis à *IEEE Trans. on Communications*, en seconde lecture depuis le 13/09/15.

2 — Synthèse des travaux

Mes travaux de recherche se situent dans le domaine du traitement du signal pour les télécommunications. Sur un plan méthodologique, les outils utilisés relèvent essentiellement de la théorie de l'optimisation, de l'algèbre et de la théorie de l'information et dans une moindre mesure de la théorie des jeux. Les thèmes traités se situent à différents niveaux de la chaîne de communication : mise en forme des signaux, correction des imperfections du canal de transmission, adaptation et optimisation du récepteur. Ils sont détaillés dans les paragraphes ci-dessous.

J'ai choisi de ne pas présenter ces travaux par ordre totalement chronologique mais de les organiser autour des thèmes suivants :

❶ Estimation adaptative de canaux de communication
❷ Codage correcteur de bruit impulsionnel
❸ Diffusion sur les canaux sans fil
❹ Techniques itératives de décodage

L'organisation choisie permet de montrer une progression entre des travaux menés en début de carrière et des travaux plus récents et plus personnels. Elle permet également de montrer des exemples de travaux transverses menés sur plusieurs années et parfois de manière discontinue (thème ❶) et aussi de montrer des exemples de travaux correspondant à des co-encadrements de thèse (thèmes ❷ et ❸). Pour chaque thème, je décrirai le problème posé, les contraintes éventuelles, la démarche adoptée ainsi que les contributions. Parmi ces quatre thèmes, j'ai choisi de mettre en avant le thème ❹. Il sera présenté dans la deuxième partie de ce document sous la forme d'un exposé scientifique détaillé. Les autres thèmes sont décrits dans ce chapitre d'une manière plus succincte. Pour garder une taille raisonnable à ce manuscrit, j'ai sélectionné un sous-ensemble représentatif de mes activités aussi certains travaux ne seront pas présentés. Je ne présenterai pas les travaux portant sur l'annulation d'écho stéréophonique réalisés lors de ma thèse, ni les travaux sur l'accélération de l'algorithme de Blahut-Arimoto, ni les travaux sur l'allocation de puissance dans les réseaux sans fil, ni les contributions concernant la diffusion sur des canaux sans fil avec contrainte de sécurité.

Dans ce chapitre et les suivants, les références aux travaux auxquels j'ai contribué apparaissent en gras et en orange. Les anglicismes, correspondant à des termes techniques ou scientifiques non traduits, sont en italique dans le texte.

2.1 Estimation adaptative de canaux de communication

Cette section est organisée autour de trois contextes différents ayant pour dénominateur commun l'estimation adaptative. C'est un sujet qui a été abondamment traité à la fin des années 90, notamment pour estimer des canaux de communication dans le contexte de la téléphonie mobile. Ce domaine applicatif a beaucoup évolué depuis, aussi j'insisterai dans ce document sur les méthodes et solutions algorithmiques plus que sur les normes (GPRS, GSM ou EDGE). Parmi les méthodes d'estimation possibles, nous retiendrons ici les méthodes dites "sous-espaces" [88, 148, 198] et les méthodes du "maximum de vraisemblance". Dans les méthodes sous-espaces, on exploite l'orthogonalité entre le sous-espace "source" et le sous-espace "bruit". En général, ces méthodes conduisent à une expression analytique de la solution ce qui est un avantage incontestable. Cet estimateur étant lié à la structure du problème, il peut se révéler inefficace en cas d'erreur de modélisation en particulier lorsque la matrice de canal est quasiment singulière. Les méthodes du type maximum de vraisemblance sont quant à elles, en général, optimales lorsque l'on dispose d'un nombre importants de données. L'estimateur peut, comme pour les méthodes sous-espace, admettre une expression analytique en particulier lorsque l'on considère le maximum de vraisemblance déterministe. Leur implémentation peut être compliquée par la présence de minima locaux. Une méthode sous-espace peut parfois être utilisée en complément afin de fournir une initialisation et éviter ainsi la convergence vers un minimum local du maximum de vraisemblance. Ce sont ces deux grandes familles d'estimateurs que nous avons considérées ici.

Les premiers travaux présentés dans la section 2.1.1 portent sur une classe d'algorithmes adaptatifs permettant l'estimation de vecteurs propres ou de sous-espace propres de matrices de covariance. Alors que la communauté traitement du signal a considéré principalement des algorithmes blocs, la communauté réseaux de neurones a, de son côté, développé de nombreux algorithmes adaptatifs. L'application de ces algorithmes au domaine du traitement du signal (apprentissage, séparation de sources, identification aveugle) a été notamment étudiée dans [20]. Le travail que nous avons conduit est une analyse asymptotique, portant sur ces algorithmes adaptatifs proposés dans la littérature des réseaux neuronaux, et aboutissant notamment à une expression analytique de l'erreur quadratique moyenne associée aux estimateurs étudiés permettant une comparaison objective des différentes alternatives proposées. Dans les thèmes exposés dans les section 2.1.2 et 2.1.3, la démarche est différente puisqu'il s'agit de construire la solution algorithmique répondant aux contraintes d'une application et/ou d'un système donné. On s'intéressera ici à l'identification conjointe canal/symbole d'une part dans un système multi-porteuse et d'autre part dans un système SIMO (*Single-Input/Multiple-Output*). Nous avons retenu cette fois une solution du type maximum de vraisemblance. Notre contribution porte sur l'analyse de convergence et la gestion du problème des minima locaux ainsi que sur la construction d'une solution adaptative à faible coût.

Ces travaux s'étendent sur une fenêtre temporelle relativement large allant de 1996 pour les travaux sur les méthodes sous-espace à 2008 pour les méthodes de type maximum de vraisemblance. Les travaux relatifs au canal multi-capteur ont commencés durant la thèse et se sont poursuivis après le recrutement. Les résultats relatifs au canal OFDM ont été obtenus lors de la thèse de Jean-Marcel Mamfoumbi soutenue en 2006.

2.1.1 Méthode sous-espace

Dans la communauté traitement du signal, les premiers travaux portant sur l'estimation de vecteurs propres ou de sous-espaces propres remontent aux années 70 et s'intéressent à des applications telles que la reconnaissance de formes ou la compression de données. Les premiers algorithmes proposés ont une complexité en $O(n^3)$ opérations, où n est la taille des vecteurs de données [1], ce qui ne permet pas de les utiliser pour l'estimation ou le suivi de quantités non-

1. Les matrices de covariance considérées sont donc de tailles $n \times n$.

stationnaires. Soit r le rang du sous-espace dominant ou minorant à estimer (avec en général $r \ll n$), le premier algorithme en $O(n^2 r)$ est proposé en 1978 dans [160], d'autres contributions [54, 108] suivent permettant de ramener la complexité de $O(n^2 r)$ à $O(nr^2)$. Au début des années 90, l'estimation adaptative de sous-espaces de matrices de covariance connaît à nouveau un regain d'intérêt puisqu'elle apparaît comme une solution possible pour l'identification aveugle de canaux de communication dans une situation multi-capteur [148, 184]. En parallèle, de nouveaux algorithmes adaptatifs sont proposés dans la littérature des réseaux de neurones [158] avec une complexité calculatoire en $O(nr)$ donnant lieu à une classe d'algorithmes à faible complexité. Au moment où nous avons abordé ce sujet, cette classe d'algorithme avait été étudiée sous deux points de vue : l'analyse de convergence en situation de pas décroissant et l'implantation neuronale [60].

Notre objectif a été l'étude et l'analyse des performances asymptotiques de chacun des estimateurs. Nous utilisons un résultat d'approximation gaussienne pour montrer que, pour un pas fixe, on peut avoir une expression littérale de la variance asymptotique de l'erreur sur les vecteurs propres ou sur la matrice de projection. Les lois asymptotiques des estimateurs sont ensuite utilisés afin d'obtenir des indices de performances (erreur quadratique moyenne, distance à l'orthonormalité) permettant d'orienter les utilisateurs vers l'algorithme présentant le meilleur compromis pour leur application. Les résultats obtenus ont été publiés dans **[61, 62]**, ils sont résumés ci-après.

Présentation des algorithmes

Notation 2.1. *Soit un vecteur aléatoire centré* $\mathbf{x}$*, gaussien et de longueur n, nous noterons* $\mathbf{R}_x = E(\mathbf{x}\mathbf{x}^T)$ *la matrice de covariance. Nous noterons* $\lambda_1 \geq \ldots \geq \lambda_n$ *les valeurs propres de* $\mathbf{R}_x$ *et* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *ses vecteurs propres associés.*

Nous nous intéressons ici d'une part à l'estimation adaptative des r vecteurs propres normalisés associés aux r plus grandes [resp. plus petites] valeurs propres distinctes $(\lambda_1, \ldots, \lambda_r)$ [resp. $\lambda_{n-r+1}, \ldots, \lambda_n$] de $\mathbf{R}_x$ et d'autre part à l'estimation d'une base orthonormée quelconque de l'espace dominant de dimension r de $\mathbf{R}_x$.

La plupart des algorithmes étudiés ont été décrits par Oja [158]. Ils peuvent tous être dérivés à partir de la méthode d'itérations simultanées d'analyse numérique [175], sous forme de l'algorithme d'approximation stochastique suivant :

$$\mathbf{W}'_{t+1} = \mathbf{W}_t + \mathbf{R}_t \mathbf{W}_t \Gamma_t \tag{2.1}$$

$$\mathbf{W}_{t+1} = \mathbf{W}'_{t+1} \mathbf{S}^{-1}_{t+1} \tag{2.2}$$

dans lequel $\mathbf{W}_t = (\mathbf{w}_{t,1}, \ldots, \mathbf{w}_{t,r}) \in \mathbb{R}^{n \times r}$ est une matrice dont les colonnes $\mathbf{w}_{t,k} \in \mathbb{R}^n$ sont orthonormées et approchent les r vecteurs propres dominants de $\mathbf{R}_x$.

Dans (2.1), la matrice Γ_t est une matrice diagonale $r \times r$ de gains vérifiant $\sum_{t=1}^{+\infty} \gamma_t = +\infty$ et $lim_{t \to +\infty} \gamma_t = 0$. Dans les algorithmes que nous étudions, nous aurons soit $\Gamma_t = \gamma_t \mathbf{I}_r$ soit $\Gamma_t = \gamma_t diag(1, \alpha_2, \ldots, \alpha_r)$ où $\alpha_i > 0$ est utilisé pour assurer un meilleur compromis entre erreur résiduelle et vitesse de convergence. La matrice $\mathbf{R}_t$ de (2.1) est une estimée de la matrice de covariance $\mathbf{R}_x$. Nous utiliserons ici, l'estimée instantanée $\mathbf{x}_t \mathbf{x}_t^T$. Dans (2.2), $\mathbf{S}_{t+1}$ est une matrice dépendant de $\mathbf{W}'_{t+1}$, qui orthonormalise les colonnes de $\mathbf{W}'_{t+1}$ conduisant à $\mathbf{W}_{t+1}$. Selon la forme de la matrice $\mathbf{S}_{t+1}$, différents algorithmes d'approximations stochastiques peuvent être dérivés [158], **[62]**.

Nous distinguerons parmi ces algorithmes cinq cas particuliers donnés dans la table 2.1. Ces cinq algorithmes ne cherchent pas tous à estimer la même chose. Le SNL est un algorithme convergeant vers une base orthonormée quelconque d'un sous espace propre dominant. Le SGA, GHA et WSA convergent vers les vecteurs propres d'un espace dominant. Le OFA converge vers les r vecteurs propres d'un espace minorant.

Analyse de performance asymptotique

Nous classons ces algorithmes en deux catégories :

Subspace Network Learning (SNL) $\mathbf{w}_{t+1,k} = \mathbf{w}_{t,k} + \gamma_t[\mathbf{I}_n - \sum_{i=1}^{r} \mathbf{w}_{t,i}\mathbf{w}_{t,i}^T]\mathbf{x}_t\mathbf{x}_t^T\mathbf{w}_{t,k} \quad k = 1,\ldots,r$
Stochastic Gradient Ascent (SGA) $\mathbf{w}_{t+1,k} = \mathbf{w}_{t,k} + \alpha_k\gamma_t[\mathbf{I}_n - \mathbf{w}_{t,k}\mathbf{w}_{t,k}^T - \sum_{i=1}^{k-1}(1+\frac{\alpha_i}{\alpha_k})\mathbf{w}_{t,i}\mathbf{w}_{t,i}^T]\mathbf{x}_t\mathbf{x}_t^T\mathbf{w}_{t,k} \quad k = 1,\ldots,r$
Generalized Hebbian Algorithm (GHA) $\mathbf{w}_{t+1,k} = \mathbf{w}_{t,k} + \gamma_t[\mathbf{I}_n - \sum_{i=1}^{k} \mathbf{w}_{t,i}\mathbf{w}_{t,i}^T]\mathbf{x}_t\mathbf{x}_t^T\mathbf{w}_{t,k} \quad k = 1,\ldots,r$
Weighted Subspace Algorithm (WSA) $\mathbf{w}_{t+1,k} = \mathbf{w}_{t,k} + \gamma_t[\mathbf{I}_n - \sum_{i=1}^{r} \frac{\beta_k}{\beta_i}\mathbf{w}_{t,i}\mathbf{w}_{t,i}^T]\mathbf{x}_t\mathbf{x}_t^T\mathbf{w}_{t,k} \quad k = 1,\ldots,r \text{ et } 0 < \beta_1 < \ldots < \beta_r$
Optimal Fitting Analyzer (OFA) $\mathbf{w}_{t+1,k} = \mathbf{w}_{t,k} + \gamma_t[\mathbf{I}_n - \mathbf{x}_t\mathbf{x}_t^T + \mathbf{w}_{t,k}\mathbf{w}_{t,k}^T\mathbf{x}_t\mathbf{x}_t^T - \mathbf{w}_{t,k}\mathbf{w}_{t,k}^T - \beta\sum_{i=k+1}^{n} \mathbf{w}_{t,i}\mathbf{w}_{t,i}^T\mathbf{x}_t\mathbf{x}_t^T]\mathbf{w}_{t,k}$ $k = n-r+1,\ldots,n$

TABLE 2.1 – Equations de remise à jour pour les 5 algorithmes étudiés

— les algorithmes stochastiques qui convergent vers des valeurs propres ($\mathbf{V} = (\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_r})$ ou $\mathbf{V} = (\mathbf{v_{n-r+1}}, \mathbf{v_{n-r+2}}, ..., \mathbf{v_n})$)

— les algorithmes stochastiques qui ne convergent que globalement vers le sous-espace propre associé à $\mathbf{V}$

Dans le premier cas, nous utiliserons un résultat général d'approximation gaussienne que nous rappelons plus loin en l'appliquant au paramètre estimé $\mathbf{V}$. Cela nous permettra d'affirmer que $\frac{1}{\sqrt{\lambda}}[\mathbf{V}(k) - \mathbf{V}]$ converge en loi (quand $\gamma \to 0$ et $k \to \infty$) vers une loi normale centrée dont on calculera la matrice de covariance.

Dans le deuxième cas, cette démarche ne peut être appliquée qu'à la matrice de projection associée $\mathbf{P} = \mathbf{V}\mathbf{V}^T$. Nous déduisons alors, de l'algorithme initial, un algorithme stochastique à pas décroissant pour $\mathbf{P}(k)$ ce qui nous permettra de donner la loi asymptotique de l'estimée de $\mathbf{P}$. Pour permettre la comparaison, ce travail sera aussi fait pour les algorithmes du premier groupe.

Nous utiliserons le résultat suivant donné dans [29, Théorème 2, p.108] dont nous rappelons les grandes lignes.

Résultat 2.1.1 — Approximation gaussienne. Soit un algorithme stochastique à pas fixe

$$\Theta_{t+1} = \Theta_t + \gamma f(\Theta_t, \mathbf{x}_t) \tag{2.3}$$

avec $\mathbf{x}_t = g(\varepsilon_t)$ où ε_t est une chaîne de Markov indépendante de Θ_t. Supposons que Θ_t converge presque sûrement vers Θ_*, seul point asymptotiquement stable de l'algorithme à pas décroissant associé. Considérons l'équation de Lyapunov

$$\mathbf{D}\mathbf{C}_\Theta + \mathbf{C}_\Theta\mathbf{D}^T + \mathbf{G} = \mathbf{O} \tag{2.4}$$

où $\mathbf{D}$ et $\mathbf{G}$ sont respectivement la dérivée du champ moyen et la covariance du champ de l'algorithme (2.3) :

$$\mathbf{D} \overset{def}{=} E[\frac{\partial f}{\partial \Theta}(\Theta, \mathbf{x}_t)]_{\Theta=\Theta_*} \tag{2.5}$$

$$\mathbf{G} \overset{def}{=} \sum_{t=-\infty}^{\infty} Cov[f(\Theta_*, \mathbf{x}_t), f(\Theta_*, \mathbf{x}_0)] \tag{2.6}$$

Si toutes les valeurs propres de $\mathbf{D}$ sont à partie réelle strictement négative, alors, en situation

stationnaire, quand $\gamma \to 0$ et $t \to \infty$, nous avons :

$$\frac{1}{\sqrt{\gamma}}(\Theta_t - \Theta_*) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \mathbf{C}_\Theta) \tag{2.7}$$

où $\mathbf{C}_\Theta$ est l'unique solution symétrique de l'équation de Lyapunov.

Ainsi Θ_t se comporte asymptotiquement (pour γ suffisamment petit) comme un estimateur gaussien, non biaisé de Θ_* de matrice de covariance $\gamma\mathbf{C}_\Theta$. Nous appliquons ce résultat général aux algorithmes SGA, GHA, WSA et OFA. Les expressions des matrices $\mathbf{D}$ et $\mathbf{G}$ pour chacun de ces algorithmes sont explicitées dans [62]. Nous en déduisons alors les théorèmes suivants [62] :

Théorème 2.1.1 Les valeurs propres de la matrice dérivée $\mathbf{D}$ du champ moyen sont réelles et strictement négatives pour les algorithmes SGA, GHA et OFA et à parties réelles strictement négatives pour l'algorithme WSA.

Le théorème 2.1.1 garantit une partie des hypothèses pour appliquer le résultat d'approximation gaussienne. Une hypothèse sous-jacente est que la solution de l'algorithme à pas décroissant converge presque sûrement vers l'unique solution asymptotiquement stable de l'ODE associée. Comme les vecteurs propres normalisés sont définis à un signe près, l'attracteur global n'est pas unique. Toutefois l'application du résultat 2.1.1 dans des situations similaires à la notre est habituellement justifiée en invoquant [29, theorem 1, p. 107]. Il reste le problème de la convergence presque sure de l'algorithme à pas décroissant qui nécessiterait des entrées bornées pour être strictement remplie. Cette condition n'est pas simple à prouver [2]. En supposant que nous pouvons tout de même utiliser le résultat 2.1.1, l'équation de Lyapounov a une solution exacte retranscrite dans le théorème suivant :

Théorème 2.1.2 Les matrices de covariance asymptotique $\mathbf{C}_W$ des algorithmes SGA, GHA, WSA et OFA qui apparaissent dans (2.7), solutions de (2.4), se mettent sous la forme commune :

$$\mathbf{C}_W = \sum_{\substack{1\le i\le r\\ 1\le k\ne i\le n}} b_{k,i}(\mathbf{e}_i^r \mathbf{e}_i^{r^T} \otimes \mathbf{v}_k \mathbf{v}_k^T) + \sum_{1\le i\ne j\le r} c_{i,j}(\mathbf{e}_i^r \mathbf{e}_j^{r^T} \otimes \mathbf{v}_j \mathbf{v}_i^T) \tag{2.8}$$

dans laquelle $(\mathbf{e}_i^r)_{i=1,\dots,r}$ désigne la base canonique de $\mathbb{R}^r$ et $\otimes$ désigne le produit de Kronecker. Pour l'algorithme OFA les sommations sont à prendre sur respectivement $n-r+1 \le i \le n, 1 \le k \ne i \le n$ et $n-r+1 \le i \ne j \le n$. Les termes $b_{i,j}$ et $c_{i,j}$ sont propres à chaque algorithme et sont donnés dans [62].

Les démonstrations des théorèmes 1 et 2 ne sont pas retranscrites ici, elles se trouvent dans l'article [62].

Dans de nombreuses applications, on est intéressé par l'estimation de la matrice de projection associée $\mathbf{P}_t \overset{def}{=} \mathbf{W}_t \mathbf{W}_t^T$. Nous avons donc travaillé sur sa distribution asymptotique. Pour cela, nous avons utilisé un théorème de continuité directement adapté de [168, Th 6.2a, p.387] que nous avons appliqué à l'application, différentiable, $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_r) \to \mathbf{P} = \sum_{k=1}^r \mathbf{w}_k \mathbf{w}_k^T$. On obtient alors

$$\frac{1}{\sqrt{\gamma}}[Vec(\mathbf{P}_t) - Vec(\mathbf{P}_*)] \xrightarrow{\mathscr{L}} \mathscr{N}(\mathbf{0}, \mathbf{C}_P) \tag{2.9}$$

lorsque $\gamma \to 0$ et $t \to +\infty$ et où $\mathbf{P}_* \overset{def}{=} \sum_{k=1}^r \mathbf{v}_k \mathbf{v}_k^T$. L'expression de $\mathbf{C}_P$ s'obtient alors selon

$$\mathbf{C}_P = \frac{dVec(\mathbf{P})}{dVec(\mathbf{W})} \mathbf{C}_W \left[\frac{d^T Vec(\mathbf{P})}{dVec(\mathbf{W})}\right]_{\mathbf{W}=\mathbf{W}^*} \tag{2.10}$$

2. voir [98] pour un exemple de preuve dans un cas particulier.

Les expressions obtenues pour chaque algorithmes sont données dans [62]. A partir de ces expressions, nous pouvons obtenir des indicateurs de performance. Nous avons considéré un indicateur simple : l'erreur quadratique moyenne entre $\mathbf{W}_t$ et $\mathbf{W}_*$ d'une part et entre $\mathbf{P}_t$ et $\mathbf{P}_*$ d'autre part. Elle peut être obtenue à partir des distributions asymptotiques de $Vec(\mathbf{W}_t)$ et de $Vec(\mathbf{P}_t)$ et en conjecturant que les convergences en loi sont accompagnées de la convergence des deux premiers moments. On obtient :

$$\mathrm{E}\|\mathbf{W}_t - \mathbf{W}_*\|_{\mathrm{Fro}}^2 = \gamma \mathrm{Tr}(\mathbf{C}_W) + o(\gamma) \quad (2.11)$$

$$\mathrm{E}\|\mathbf{P}_t - \mathbf{P}_*\|_{\mathrm{Fro}}^2 = \gamma \mathrm{Tr}(\mathbf{C}_P) + o(\gamma) \quad (2.12)$$

Notre étude suppose que le pas γ est suffisamment proche de 0. Nous proposons ici de quantifier la zone de validité du résultat théorique. Pour cela, nous représentons sur la figure 2.1 pour chacun des algorithmes le rapport de l'erreur quadratique moyenne (une fois la convergence atteinte) sur l'erreur quadratique moyenne théorique.

Conditions expérimentales 2.1 Les paramètres de la simulation sont $n = 4$, $r = 2$ associés avec la matrice de covariance $\mathbf{R}_x = Diag(1.75, 1.5, 0.5, 0.25)$. Les vecteurs initiaux $\mathbf{w}_{0,k}$ sont choisis aléatoirement suivant une loi uniforme dans $[0;1]$ puis normalisés. On a considéré ici un SGA avec $\alpha_2 = 2$, un WSA avec $\frac{\beta_2}{\beta_1} = 0.9$ et un OFA avec $\beta = 5$.

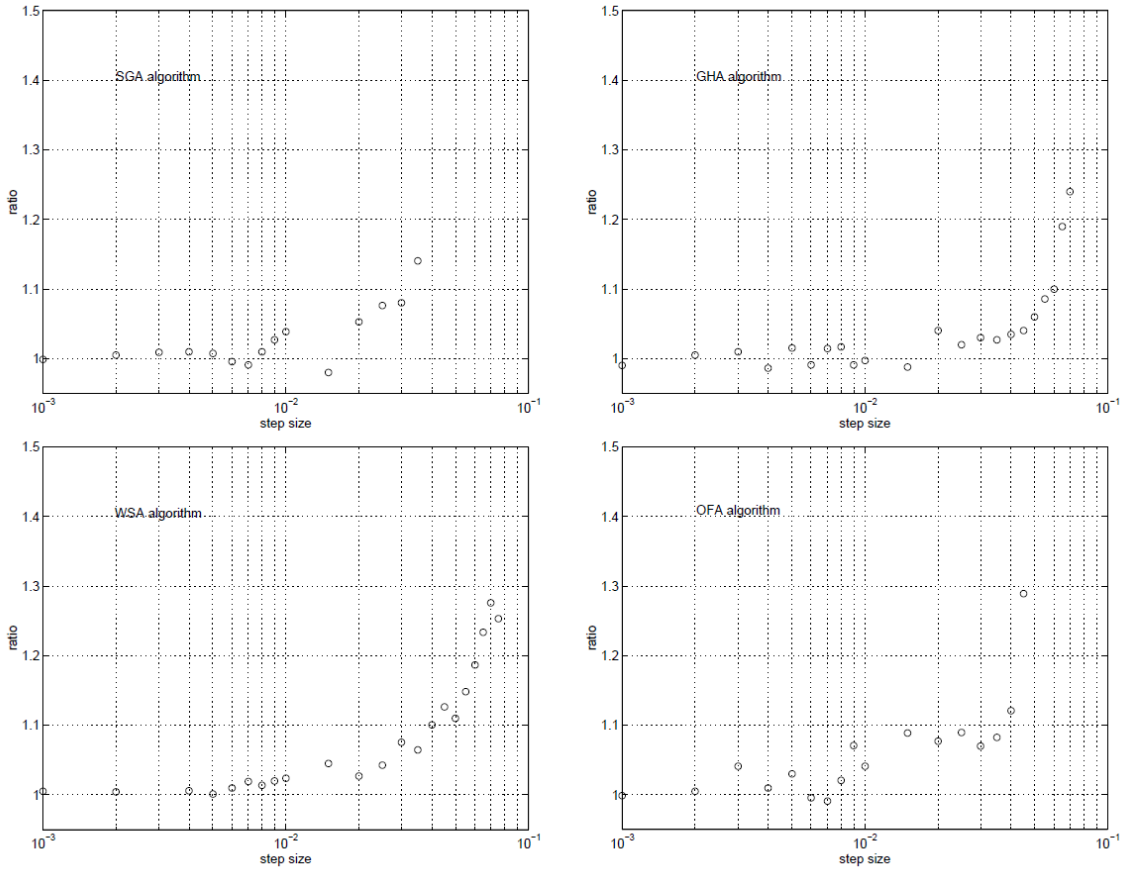


FIGURE 2.1 – Ratio de l'EQM estimée $\mathrm{E}\|\mathbf{P}_t - \mathbf{P}_*\|_{\mathrm{Fro}}^2$ (en moyennant 400 réalisations) sur l'EQM théorique $\gamma \mathrm{Tr}(\mathbf{C}_P)$ en fonction de γ.

On constate que notre analyse asymptotique est valide pour une grande gamme de valeurs de γ ($\gamma \leq 0.01$) et que le domaine de stabilité pour lequel γ reste proche de 1 est $\gamma \leq 0.035$. La déviation

à l'orthonormalité $d^2(\gamma) = \mathrm{E}\| \mathbf{W}_t^T \mathbf{W}_t - \mathbf{I}_r \|_{\mathrm{Fro}}^2$ est représentée sur la figure 2.2. On constate qu'elle est proportionnelle à γ (GHA,OFA) ou à γ^2 (WSA, SGA,Yang) dans le domaine de validité. Dans

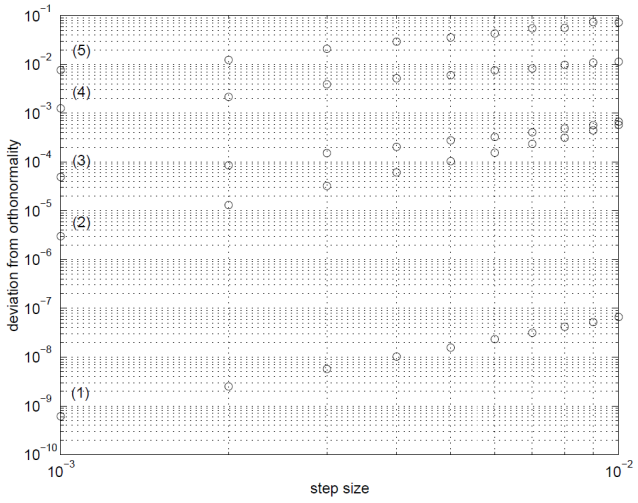


FIGURE 2.2 – Distance à l'orthogonalité $d^2(\gamma) \stackrel{def}{=} \mathrm{E}\| \mathbf{W}_t^T \mathbf{W}_t - \mathbf{I}_r \|_{\mathrm{Fro}}^2$ à la "convergence" estimée en moyennant 100 réalisations pour les algorithmes de Yang (1) [63], SGA (2), WSA (3), GSA (4), OFA (5).

cette section, nous avons étudié le comportement asymptotique d'un certain nombre d'algorithmes afin de permettre un choix judicieux en fonction de l'application tant sur le choix de l'algorithme que de ces paramètres. Dans la suite de cette section, nous allons définir un contexte applicatif particulier ainsi que les contraintes et hypothèses qui en découlent et nous allons proposer deux solutions algorithmiques dédiées.

2.1.2 Système multi-capteur

Nous considérons maintenant l'identification et l'égalisation aveugle multi-canaux. Elle a beaucoup été étudiée à la fin des années 90 pour la téléphonie mobile et c'est également dans ce cadre que les travaux présentés ici ont été menés. Les méthodes aveugles sont utiles lorsque les séquences d'apprentissage ne sont pas présentes ou exploitables. Elles sont également compatibles avec l'utilisation de séquences d'apprentissage qui fournissent un point d'initialisation alors que l'algorithme d'identification aura pour rôle de suivre les éventuelles variations du canal. Les premiers travaux en égalisation aveugle reposent sur l'estimation des statistiques d'ordre supérieur. Le recours aux seules statistiques d'ordre 2 remonte aux travaux de Tong *et al.* [184] qui montrent qu'elles sont suffisantes à condition d'introduire de la diversité spatiale ou temporelle dans le système et tant que les sous-canaux n'ont pas de zéros en commun. On utilisera dans tout ce qui suit l'acronyme SIMO (*Single-Input Multiple-Output*) pour nommer un système avec diversité spatiale ou temporelle en sortie.

Pour des canaux à variations rapides, les statistiques du signal d'entrée peuvent être inconnues ou difficiles à estimer, on peut alors avoir recours à des méthodes dites déterministes. C'est ce choix que nous avons fait ici. Nous considérons comme point de départ le maximum de vraisemblance déterministe (MVD) ; c'est un estimateur efficace à haut rapport signal à bruit. Dans une approche déterministe, les symboles et le canal sont considérés comme des inconnues. L'optimisation conjointe de la fonction de vraisemblance en le canal et les symboles est un problème difficile en général que nous pouvons résoudre ici car l'observation est une fonction linéaire du canal mais aussi des symboles. Le TSML [100] et l'IQML [181] sont deux contributions majeures du type MVD. Ces

algorithmes sont itératifs et procèdent en deux étapes, chacune ayant pour but d'identifier le canal uniquement, les symboles étant estimés une fois le canal obtenu ce qui ne permet pas d'introduire de l'information *a priori* sur les symboles. Or nous avons à notre disposition de l'information sur les symboles puisque nous connaissons la modulation choisie. Afin de pouvoir l'exploiter, nous avons préféré suivre l'approche [80] qui obtient un algorithme itératif également en deux étapes quadratiques mais qui résout à chaque étape un problème des moindres carrés alternativement par rapport au canal puis par rapport aux symboles.

Notre contribution tient dans les points suivants : l'incorporation d'un *a priori* sur les symboles dans le critère, l'étude de l'influence de l'*a priori* sur les minima locaux et enfin la construction d'une solution adaptative rapide, à faible complexité et capable de suivre les variations du canal tout en étant robuste aux erreurs de modélisation telle que la sur-estimation de l'ordre du canal. Les résultats obtenus ont été publiés dans **[15, 18]** et dans plusieurs conférences internationales.

En collaboration avec la société Wavecom, nous avons considéré l'application de cette méthode aux communications de type GSM à haute vitesse. Nous avons écrit une version semi-aveugle de l'algorithme puisque des symboles d'apprentissage sont disponibles. Nous avons ensuite apporté des améliorations et aménagements à l'algorithme afin de le rendre compétitif avec l'algorithme de Viterbi à faible et moyenne vitesse et plus performant dans les situations à très haute vitesse. Cette deuxième partie a donné lieu à un brevet et à une communication à ICASSP'04.

Modèle et critère d'optimisation

Nous considérons donc ici un système SIMO ayant pour entrée $\widetilde{s}(n)$, une séquence de symboles binaires, et $x_i(k)$ comme sortie numéro i avec $1 \leq i \leq L$. L'observation $x_i(n)$ est modélisée comme le résultat du filtrage de la séquence $\widetilde{s}(n)$ par le filtre inconnu $\widetilde{\mathbf{h}}_i$, $1 \leq i \leq L$ auquel s'ajoute un bruit $b_i(n)$. On a

$$x_i(n) = \sum_{k=0}^{k=M} \widetilde{s}(n-k)\widetilde{h}_i(k) + b_i(n) \;\; i = 1,...,L \tag{2.13}$$

où $\{b_i(n)\}$, $1 \leq i \leq L$ est un bruit additif gaussien *iid* de moyenne nulle, de variance σ_b^2 et mutuellement décorrélés.

Notation 2.2. *On notera :*

- *M l'ordre maximum des canaux*
- $\mathbf{h}$ *et* $\mathbf{s}_N$ *un canal et un vecteur de symboles.*
- $\widetilde{\mathbf{h}}$ *et* $\widetilde{\mathbf{s}_N}$ *les vrais canaux et symboles transmis.*
- $\widehat{\mathbf{h}}$ *et* $\widehat{\mathbf{s}_N}$ *les estimés de* $\widetilde{\mathbf{h}}$ *et* $\widetilde{\mathbf{s}_N}$.
- $\mathbf{s}_N(n) = [s(n), s(n-1), ...s(n-N-M+1)]^T$ *un vecteur de longueur* $N+M$ *et n un index temporel.*

L'équation (2.13) peut s'écrire à l'aide de notations vectorielles. Soit $\mathbf{X}_N(n) = [x_1(n)...x_L(n)....x_1(n-N+1)...x_L(n-N+1)]^T$ le vecteur d'observation obtenu en entrelaçant les sorties des différents filtres, on construit de même le vecteur de bruit $\mathbf{B}_N(n)$ ce qui conduit à :

$$\mathbf{X}_N(n) = \mathscr{T}_N(\widetilde{\mathbf{h}})\widetilde{\mathbf{s}}_N(n) + \mathbf{B}_N(n) \tag{2.14}$$

où $\widetilde{\mathbf{h}}(k) = [\widetilde{h}_1(k)...\widetilde{h}_L(k)]^T$. L'opérateur $\mathscr{T}_N$ introduit dans (2.14) transforme toute suite de vecteurs $\mathbf{h}(k) = [h_1(k)...h_L(k)]^T$ en une matrice de Sylvester de taille $LN \times M+N$ selon :

$$\mathscr{T}_N(\mathbf{h}) = \begin{pmatrix} \mathbf{h}(0) & \dots & \mathbf{h}(M) & 0 & \dots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{h}(0) & \dots & \mathbf{h}(M) \end{pmatrix}$$

De manière duale, $\mathscr{U}$ transforme un vecteur $\mathbf{s}_N$ en une matrice $\mathscr{U}(\mathbf{s}_N(n))$ de taille $LN \times L(M+1)$ telle que :

$$\mathscr{U}(\mathbf{s}_N(n))\mathbf{h} = \mathscr{T}_N(\mathbf{h})\mathbf{s}_N(n), \quad \forall \mathbf{s}_N, \forall \mathbf{h} \tag{2.15}$$

On montre aisément que cet opérateur est de la forme :

$$\mathscr{U}(\mathbf{s}_N(n)) = \begin{pmatrix} I_L \bigotimes s_1(n)^T \\ I_L \bigotimes s_1(n-1)^T \\ \vdots \\ I_L \bigotimes s_1(n-N+1)^T \end{pmatrix}$$

où $\bigotimes$ symbolise le produit de Kronecker et $\mathbf{I}_L$ est la matrice identité de taille $L \times L$. Les résultats obtenus nécessitent les hypothèses usuelles suivantes :

- H_1 $\mathscr{T}_N(\widetilde{\mathbf{h}})$ est de rang colonne plein.
- H_2 La complexité linéaire [3] de $\widetilde{\mathbf{s}_N}$ est au moins égale à $2M+1$ [36, 85].
- H_3 L'ordre maximum des canaux est connu ou correctement estimé.
- H_4 Les symboles appartiennent à une modulation de type PSK (*Phase Shift Keeying*).

Le critère considéré est le maximum de vraisemblance déterministe donné par

$$\mathbb{J}(\mathbf{h}, \mathbf{s}_N(n)) = \|\mathbf{X}(n) - \mathscr{T}_N(\mathbf{h})\mathbf{s}_N(n)\|^2 \tag{2.16}$$

$$= \|\mathbf{X}_N(n) - \mathscr{U}(\mathbf{s}_N(n))\mathbf{h}\|^2 \tag{2.17}$$

Dans le cas non bruité, on sait que le minimum global est atteint uniquement pour le vrai canal et les vrais symboles à un facteur d'échelle près [82, 164]. Un critère déterministe est en général préféré dans les situations à haute vitesse pour lesquelles il est difficile d'obtenir de bonnes estimations statistiques. Le critère (2.16) répond à ce problème. D'un autre côté si le système est à variation lente et si on dispose de suffisamment de données, on doit privilégier une solution statistique qui conduira *in fine* à une meilleure estimation. Afin de construire un algorithme compatible avec un canal à variation lente ou à variation rapide, nous proposons ici une solution intermédiaire entre méthode statistique et déterministe. Pour cela, nous considérons que les symboles ne sont plus des quantités déterministes mais des variables aléatoires qui obéissent à une certaine loi statistique. Nous savons que les symboles appartiennent à un alphabet fini. On sait que l'introduction d'un *a priori* aussi fort (appartenance à un ensemble discret de faible cardinalité) entraîne en général une augmentation significative du nombre de minima locaux. Pour pallier à cela, nous considérons l'introduction d'un *a priori* partiel, correspondant à la loi de probabilité suivante :

$$\begin{cases} p(s(k)) = 0 & \text{si} \quad \|s(k)\| > 1 \\ p(s(k)) = \frac{1}{Z} e^{\kappa \|s(k)\|^2} & \text{si} \quad \|s(k)\| \leq 1 \\ p(\mathbf{s}) = \Pi_k p(s(k)) \end{cases} \tag{2.18}$$

où Z est une constante de normalisation et où la valeur de κ permet de régler la "quantité" d'*a priori* introduite dans le système. En particulier choisir $\kappa = 0$ conduit à un *a priori* uniforme sur la boule unité alors que $\kappa \to \infty$ conduit à un *a priori* uniforme sur la sphère unité. L'algorithme obtenu a un coût de calcul plus faible qu'une méthode du type maximum de vraisemblance statistique tout en améliorant la qualité de l'estimation comparativement à la méthode déterministe. Le critère devient alors

$$\mathscr{L}_\gamma(\mathbf{h}, \mathbf{s}_N) = \|\mathbf{X}(n) - \mathscr{T}_N(\mathbf{h})\mathbf{s}_N\|^2 - \gamma\|\mathbf{s}_N\|^2 \quad \mathbf{h} \in \mathbb{C}^{L(M+1)} \quad \mathbf{s}_N \in \mathscr{E}^{M+N} \tag{2.19}$$

3. La complexité linéaire mesure la prédictibilité d'une séquence déterministe de longueur finie. De manière plus formelle, la complexité linéaire de $\{\widetilde{s}(n-k)\}_{k=0}^{k=N+M-1}$ est définie comme la plus petite valeur c telle qu'il existe $\{\lambda_j\}_{j=1}^{c}$ où $\widetilde{s}(n-i) = -\sum_{j=1}^{c} \lambda_j \widetilde{s}(n-i-j)$ avec $i = c, \ldots, N+M-1$. Si la complexité linéaire de $\widetilde{\mathbf{s}_N}$ est au moins égale à $2M+1$ alors $\mathscr{U}(\mathbf{s}_N(n))$ est de rang colonne plein.

où $\gamma = 2\sigma_b^2\kappa$ et $\mathscr{E}^{M+N} = \{\mathbf{s}_N \in \mathbb{C}^{M+N} : \ \|s(k)\| \leq 1, k = 0...M+N-1\}$. Le critère $\mathscr{L}_\gamma(\mathbf{h}, \mathbf{s}_N)$ est convexe par rapport à chaque variable prise séparément tant que $\gamma \leq \lambda_{min}$ où λ_{min} est la plus petite valeur propre de $\mathscr{T}_N(\mathbf{h})^H \mathscr{T}_N(\mathbf{h})$. Dans le cas non-convexe, le problème de programmation quadratique soumis à des contraintes linéaires est NP-complet [150]. La recherche de minima locaux a alors une complexité calculatoire qui augmente exponentiellement avec le nombre de variables [26, 40, 99, 162]. Aussi, dans toute la suite nous considérerons que $\gamma \leq \lambda_{min}$. L'estimation est ensuite réalisée par une procédure itérative qui alterne les deux minimisations suivantes :

$$\widehat{\mathbf{h}}^{(k)} = [\mathscr{U}(\widehat{\mathbf{s}}_N^{(k-1)})^H \mathscr{U}(\widehat{\mathbf{s}}_N^{(k-1)})]^{-1} \mathscr{U}(\widehat{\mathbf{s}}_N^{(k-1)})^H \mathbf{X}_N(n) \tag{2.20}$$

$$\widehat{\mathbf{s}}_N^{(k)} = \arg \min_{\mathbf{s}_N \in \mathscr{E}^{M+N}} \mathscr{L}_\gamma(\widehat{\mathbf{h}}^{(k)}, \mathbf{s}_N) \tag{2.21}$$

Chaque étape diminue la valeur du critère donc l'algorithme converge, éventuellement vers un minimum local. Le problème d'optimisation dans (2.21) est résolu par une méthode de relaxation [51], les équations de remise à jour sont données dans [18]. Cette classe d'algorithme sera nommée CMLBA$_\gamma$ (Conditionnal Maximum Likelihood Block Algorithm) pour le distinguer des versions adaptatives proposées ensuite. Le paramètre γ peut être choisi dans un intervalle allant de 0 jusqu'à λ_{min}. Nous avons étudié la convergence du CMLBA$_\gamma$ et dans un premier temps l'unicité du minimum global.

Théorème 2.1.3 Si $\mathbf{B}_N = \mathbf{0}$ et si les hypothèses 1 et 2 sont vérifiées, le minimum global de $\mathscr{L}_\gamma(\mathbf{h}, \mathbf{s}_N)$ sur $\mathbb{C}^{L(M+1)} \times \mathscr{E}^{M+N}$ est $(\alpha\widetilde{\mathbf{h}}, \frac{\widetilde{\mathbf{s}}_N(n)}{\alpha})$. Si $\gamma > 0$ alors $\|\alpha\| = 1$.

Pour $\gamma > 0$ et en l'absence de bruit, le minimum global conduit à l'estimation du canal et des symboles émis à un déphasage près. Les relations de Kuhn-Tucker [51] donnent une condition nécessaire et suffisante pour caractériser les points stationnaires de l'algorithme. Ils doivent satisfaire :

$$\widehat{\mathbf{h}} = [\mathscr{U}(\widehat{\mathbf{s}}_N)^H \mathscr{U}(\widehat{\mathbf{s}}_N)]^{-1} \mathscr{U}(\widehat{\mathbf{s}}_N)^H \mathbf{X}_N(n); \quad \widehat{\mathbf{s}}_N \in \mathscr{E}_{M+N} \tag{2.22}$$

$$\widehat{\mathbf{s}}_N(n) = \left[\mathscr{T}_N(\widehat{\mathbf{h}})^H \mathscr{T}_N(\widehat{\mathbf{h}}) - \gamma \mathbf{I}_{M+N} + \Phi\right]^{-1} \mathscr{T}_N(\widehat{\mathbf{h}})^H \mathbf{X}_N(n); \quad \widehat{\mathbf{s}}_N \in \mathscr{E}_{M+N} \tag{2.23}$$

$$\Phi = diag(\phi_i)_{0 \leq i \leq M+N-1}, \tag{2.24}$$

$$\phi_i \geq 0 \text{ et } \phi_i(\|\widehat{s}(n-i)\|^2 - 1) = 0 \ \forall \ i = 0, ..., M+N-1 \tag{2.25}$$

La différence entre les points stationnaires de l'algorithme sans *a priori* et sans contrainte et du CMLBA$_\gamma$ est le terme $\Phi - \gamma \mathbf{I}_{M+N}$ apparaissant dans la $2^{ème}$ équation. La matrice Φ est liée a l'*a priori* de la manière suivante :

Propriété 2.1.1 Soit $(\widehat{\mathbf{h}}, \widehat{\mathbf{s}}_N(n))$ un point stationnaire du CMLBA$_\gamma$ et Φ la matrice définie dans (2.22-2.25) alors $trace(\Phi) = \gamma \|\widehat{\mathbf{s}}_N\|^2$.

A partir de cette proposition, on peut montrer que lorsque $\gamma > 0$, il existe $i_0 \in \{0, ..., M+N-1\}$ tel que $\phi_{i_0} > 0$ et par conséquent $\|\widehat{s}(n-i-0)\|^2 = 1$. Le paramètre γ a donc pour effet de repousser les symboles estimés vers la frontière de l'ensemble $\mathscr{E}_{M+N}$. Lorsque $\gamma = 0$, la seule différence avec le critère (2.16) est la contrainte faite aux symboles d'être dans $\mathscr{E}_{M+N}$. On peut montrer [18] que l'ajout de cette contrainte n'a pas d'influence sur le nombre de minima locaux [4] bien qu'elle ait un effet positif sur les performances.

Dans un second temps, nous avons porté notre attention sur le cas où la matrice $\mathscr{T}_N(\mathbf{h})^H \mathscr{T}_N(\mathbf{h})$ est mal conditionnée et présente une valeur propre, λ_{min}, proche de 0. C'est une situation que l'on rencontre lorsque les réponses impulsionnelles des canaux sont similaires. Vis-à-vis de l'algorithme

4. Si $(\widehat{\mathbf{h}}, \widehat{\mathbf{s}_N}(n))$ est un point stationnaire de l'algorithme non contraint alors $(\alpha\widehat{\mathbf{h}}, \frac{\widehat{\mathbf{s}_N}(n)}{\alpha})$ est aussi un point stationnaire et $(\alpha\widehat{\mathbf{h}}, \frac{\widehat{\mathbf{s}_N}(n)}{\alpha})$ est un point stationnaire de l'algorithme avec contrainte pour les valeurs de α telles que $\frac{\widehat{\mathbf{s}_N}(n)}{\alpha} \in \mathscr{E}_{M+N}$.

proposé, l'*a priori* se trouve limité par l'intermédiaire de γ qui est contraint entre 0 et λ_{min}. Lorsque λ_{min} est proche de 0 seul un *a priori* faible peut être introduit ce qui peut s'avérer insuffisant dans une situation déjà pathologique. Pour y remédier, nous avons proposé une solution algorithmique dérivée du théorème ci-dessous.

Théorème 2.1.4 Soit $\mathbf{A_1}$ et $\mathbf{A_2}$ deux matrices telles que $\mathbf{A} = [\mathbf{A_1}\ \ \mathbf{A_2}]$, on a :

$$\lambda_{min}^{A_1} \geq \lambda_{min}^{A} \quad \text{et} \quad \lambda_{min}^{A_2} \geq \lambda_{min}^{A}$$

oú λ_{min}^{A} resp. $\lambda_{min}^{A_i}$ est la plus petite valeur propre de $\mathbf{A}^H\mathbf{A}$ resp. $\mathbf{A_i}^H\mathbf{A_i}$, i=1,2.

Cela signifie que si l'on partitionne la matrice de canal [5], le conditionnement de la matrice courante sera au moins égal à celui de la matrice de départ. On peut alors espérer pouvoir introduire dans chacun des sous-problème ainsi créé un *a priori* plus fort que si l'on considérait la matrice de canal dans son ensemble. Les nouvelles équations de mise à jour sont données ci-dessous dans le cas de 2 partitions.

$$\widehat{\mathbf{h}}^{(k)} = [\mathscr{U}(\widehat{\mathbf{s}}_N^{(k-1)})^H \mathscr{U}(\widehat{\mathbf{s}}_N^{(k-1)})]^{-1} \mathscr{U}(\widehat{\mathbf{s}}_N^{(k-1)})^H \mathbf{X}_N(n) \tag{2.26}$$

$$\widehat{\mathbf{u}}^{(k)} = \arg\min_{\mathbf{u}\in\mathscr{E}^{N_1}} \mathscr{L}_{\gamma_1}\left(\widehat{\mathbf{h}}^{(k)}, \begin{bmatrix} \mathbf{u} \\ \widehat{\mathbf{v}}^{(k-1)} \end{bmatrix}\right) \tag{2.27}$$

$$\widehat{\mathbf{v}}^{(k)} = \arg\min_{\mathbf{v}\in\mathscr{E}^{N_2}} \mathscr{L}_{\gamma_2}\left(\widehat{\mathbf{h}}^{(k)}, \begin{bmatrix} \widehat{\mathbf{u}}^{(k)} \\ \mathbf{v} \end{bmatrix}\right) \tag{2.28}$$

$$\widehat{\mathbf{s}}_N^{(k)} = [(\widehat{\mathbf{u}}^{(k)})^H (\widehat{\mathbf{v}}^{(k)})^H]^H \tag{2.29}$$

avec $\gamma_1 = \lambda_{min}(\mathscr{T}_N^{1\to N_1}(\widehat{\mathbf{h}}))^H \mathscr{T}_N^{1\to N_1}(\widehat{\mathbf{h}}))$, $\gamma_2 = \lambda_{min}(\mathscr{T}_N^{N_1\to N_1+N_2}(\widehat{\mathbf{h}}))^H \mathscr{T}_N^{N_1\to N_1+N_2}(\widehat{\mathbf{h}}))$ et $\mathscr{T}_N(\mathbf{h}) = [\mathscr{T}_N^{1\to N_1}(\mathbf{h})\ \ \mathscr{T}_N^{N_1\to N_1+N_2}(\mathbf{h})]$. La généralisation à P partitions ne pose pas de problème sachant que tous les blocs ont des longueurs identiques. Nous nommerons CMLBA$_\gamma(P)$ la procédure ainsi construite avec P partitions.

Vers un algorithme adaptatif

Le CMLBA$_\gamma$ peut être transformé en algorithme adaptatif en introduisant un facteur d'oubli $\lambda \in [0\ \ 1]$ dans le critère (2.19) selon :

$$\mathscr{L}_\lambda^{(W)}(\mathbf{h}, \mathbf{s}_N) = \|\Lambda_N^{1/2}[\mathbf{X}_N(n+k) - \mathscr{T}_N(\mathbf{h})\mathbf{s}_N]\|^2 - \gamma\|\mathbf{s}_N\|^2 \tag{2.30}$$

avec $\Lambda_N = diag([1,...,1,\lambda,...,\lambda,....,\lambda^{N-1}...\lambda^{N-1}])$ et où Λ_N est de taille $LN \times LN$. Le facteur d'oubli ne s'applique pas à l'*a priori* qui est indépendant du temps. On applique ensuite les étapes suivantes. On recherche le minimum de (2.30) par rapport à chaque variable prise séparément :

$$\widehat{\mathbf{s}}_{N+k}^{(k)}(n+k) = \arg\min_{\mathbf{s}_{N+k}\in\mathscr{E}^{M+N+k}} \mathscr{L}_\gamma^{(W)}\left(\widehat{\mathbf{h}}^{(k-1)}, \mathbf{s}_{N+k}\right) \tag{2.31}$$

$$\widehat{\mathbf{h}}^{(k)} = \arg\min_{\mathbf{h}} \mathscr{L}_\gamma^{(W)}\left(\mathbf{h}, \widehat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right) \tag{2.32}$$

Dans l'équation (2.31), on remet à jour $M+N+k$ symboles. La complexité calculatoire augmente donc avec k. Nous proposons de calculer, à chaque itération le nouveau symbole émis et nous remettons à jour uniquement les K précédents. Une étude par simulation montre que choisir K égal à l'ordre du canal conduit à un bon compromis. Le canal estimé $\widehat{\mathbf{h}}^{(k)}$ est remis à jour récursivement à partir de $\widehat{\mathbf{h}}^{(k-1)}$. Cette opération se fait sans approximation à l'aide d'un RLS ou d'un LMS. La

5. Cela revient à partitionner le vecteur de symboles et à résoudre consécutivement le problème de minimisation par rapport à chaque morceau du vecteur.

matrice de covariance qui intervient dans le RLS est la matrice de covariance des symboles, elle est diagonale sous nos hypothèses. Remplacer le RLS par un LMS a donc très peu d'incidence sur les performances.

L'algorithme résultant est noté CMLAA-γ. Il est très proche d'un algorithme de type DFE (Decision Feedback Equalizer) pour lequel on aurait remplacé la décision dure par des estimées souples de tous les symboles présents dans la ligne à retard du filtre [15].

Pour une application en téléphonie mobile, le CMLAA-γ a un certain nombre de points positifs. L'ajout de l'*a priori* permet d'augmenter la vitesse de convergence et les capacités de poursuite de l'algorithme. La complexité calculatoire est limitée grâce à l'utilisation d'un LMS pour la remise à jour du canal et grâce à une remise à jour des symboles les plus récents (même ordre de grandeur que l'ordre du canal). On pourra voir aussi à partir des résultats numériques que cet algorithme est robuste à la sur-estimation de l'ordre du canal ce qui est requis pour que l'algorithme ait une utilité pratique en téléphonie mobile. Une étude de convergence du CMLAA-γ a été menée dans [15] donnant les conditions pour lesquelles la convergence vers un minimum global est assurée.

Simulations

Sont montrés ci-après quelques résultats de simulation, des résultats plus complets son donnés dans [15, 18]. Les performances des algorithmes sont mesurées par l'erreur quadratique moyenne normalisée exprimée en dB :

$$NRMSE_{dB}(\mathbf{h}) = 20log_{10}\left(\frac{1}{\|\widetilde{\mathbf{h}}\|}\sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r}\|\widehat{\alpha}^{(i)}\widehat{\mathbf{h}}^{(i)}-\widetilde{\mathbf{h}}\|^2}\right)$$

où $\widehat{\mathbf{h}}^{(i)}$ est le canal estimé lors de la $i^{ème}$ réalisation et $\widehat{\alpha}^{(i)} = \arg\min_\alpha \|\alpha\widehat{\mathbf{h}}^{(i)}-\widetilde{\mathbf{h}}\|^2$. Les algorithmes suivants sont comparés :

- MLBA : algorithme bloc sans l'introduction ni de la contrainte ni de l'*a priori* [80].
- CMLBAγ : nous considérerons les deux extrêmes, le CMLBA$_{min}$ correspondant à $\gamma = 0$ et le CMLBA$_{max}$ correspondant à $\gamma = \lambda_{min}(\mathscr{T}_N(\widehat{\mathbf{h}})^H\mathscr{T}_N(\widehat{\mathbf{h}}))$.
- CMLBA$\gamma(N_p)$: algorithme CMLBAγ avec N_p partitions du vecteur de symboles.
- MLPA : algorithme de prédiction linéaire multi-pas proposé dans [81] et indépendamment dans [65, 66]. Il exploite complètement la structure du canal ce qui lui permet d'avoir de meilleures performances que les algorithmes de prédiction classiques [7, 180].
- TSML : c'est une référence parmi les algorithmes de type maximum de vraisemblance déterministe. Il a été proposé par Hua dans [100].
- CMLAAγ : version adaptative du CMLBAγ.

Conditions expérimentales 2.2 On teste les algorithmes sur deux types de canaux différents. Le canal $\mathbf{h}^{\mathbf{Hua}}$ ci-dessous :

$$\widetilde{\mathbf{h}}_1^{\mathbf{Hua}}(z) = (1-e^{j\theta_1}z^{-1})(1-e^{-j\theta_1}z^{-1}) \qquad \widetilde{\mathbf{h}}_2^{\mathbf{Hua}}(z) = (1-e^{j(\theta_1+\delta)}z^{-1})(1-e^{-j(\theta_1+\delta)}z^{-1}) \quad (2.33)$$

où θ_1 et $\theta_1+\delta$ représentent les positions angulaires des zéros sur le cercle unité et où δ est la distance entre les deux zéros du canal. Nous pouvons ainsi évaluer la robustesse des solutions proposées vis à vis de la diversité du canal. Ce canal a également été utilisé dans [100] pour comparaison des performances du TSML vis à vis de l'algorithme sous-espace proposé dans [148] ainsi que de l'algorithme *Cross-Relation* [123]. Ces algorithmes ont également été comparés à la borne de Cramer-Rao. Le second canal considéré est un canal de propagation à 12 trajets simulé en utilisant le modèle de Clarke et dont les caractéristiques sont données dans la table 2.2.

Les algorithmes blocs sont comparés sur la figure 2.3. Le canal considéré est $\mathbf{h}^{\mathbf{Hua}}$. Nous traçons le $NRMSE_{dB}$ en fonction de δ (distance entre les zéros du canal). Le rapport signal à bruit vaut 45dB. Nous observons en particuliers que le CMLBA$_{min}$ est plus performant que le TSML et

Trajet	1	2	3	4	5	6	7	8	9	10	11	12
Délai (μs)	0	0.2	0.4	0.6	0.8	1.2	1.4	1.8	2.4	3	3.2	5
Atténuation (dB)	-4	-3	0	-2	-3	-5	-7	-5	-6	-9	-11	-10

TABLE 2.2 – Profil des trajets pour le canal test (modèle COST-GSM)

que le MLBA pour des faibles valeurs de δ pourtant le CMLBA$_{min}$ a les mêmes minima locaux que le MLBA. Le CMLBA$_{max}$ a un $NRMSE_{dB}$ inférieur de 5dB à celui du CMLBA$_{min}$ pour δ suffisamment grand. Nous observons sans surprise que lorsque δ tend vers 0, les performances du CMLBA$_{max}$ se rapprochent de celles du CMLBA$_{min}$. Le CMLBA$_{max}(5)$ a été construit pour avoir de bonnes performances quel que soit le conditionnement de la matrice de canal. Nous observons que le $NRMSE_{dB}$ vaut -55dB pour toutes les valeurs de δ.

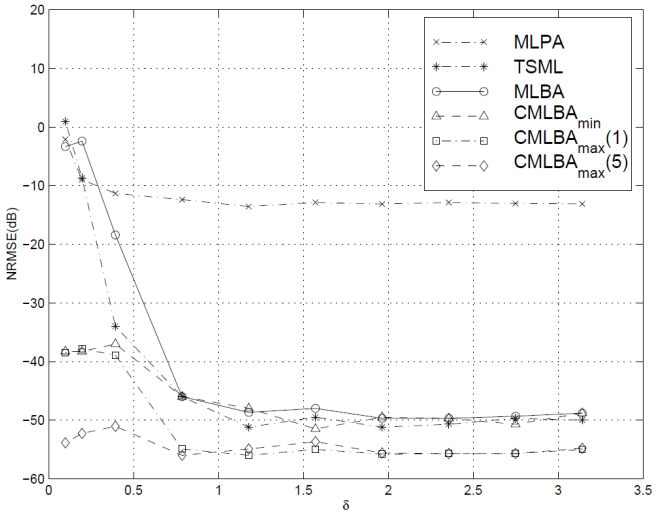


FIGURE 2.3 – Comparaison des algorithmes bloc en fonction de δ (100 réalisations, $M+N=32$, modulation BPSK).

La robustesse du CMLAA$_\gamma$ à la surestimation de l'ordre du canal est évalué dans la figure 2.4. Nous considérons cette fois le canal de la table 2.2. L'ordre du vrai canal est $\widetilde{M}=3$ et il y a deux sous-canaux. Une modulation BPSK est utilisée. Le rapport signal à bruit vaut 10dB. Nous comparons le CMLAA$_{min}$, le CMLAA$_{max}$ au RLS supervisé. L'ordre estimé du canal vaut $\widehat{M}=3$ (figure de gauche) puis $\widehat{M}=4$ (figure de droite). Les deux algorithmes sont robustes à la surestimation de l'ordre du canal.

Enfin, la capacité de poursuite du CMLAA$_\gamma$ est illustrée sur la figure 2.5. Nous nous intéressons aux comportement du CMLAA$_\gamma$ lorsque le canal subit un changement brutal [205]. Le canal initial est $\mathbf{h}^{\mathbf{Hua}}$ avec $\delta=\pi$ et $\theta_1=\pi/10$. L'ordre du canal et ses paramètres changent à l'instant $n=151$ où deux zéros $Z_1=1$ et $Z_2=-1$ sont ajoutés à chaque sous-canal. L'ordre estimé est $\widehat{M}=3$ tout au long de la simulation. La figure (2.5) montre la capacité du CMLAA$_{min}$ et du CMLAA$_{max}$ à suivre les variations du canal.

Application au GSM

Nous présentons ici les travaux entrepris en collaboration avec la société Wavecom ayant pour objectif de trouver une solution d'égalisation/identification de canaux pour le GSM dans des situations à haute-vitesse ($\approx$ 300km/h). Nous commençons par présenter les spécificités liées aux communications GSM. Ce sont des communications dans lesquelles les données sont découpées en

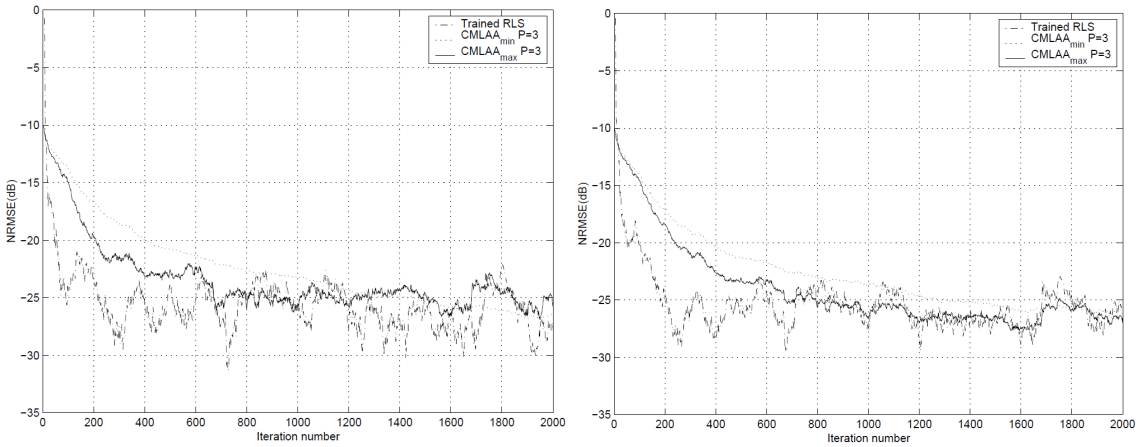


FIGURE 2.4 – NRMSE(**h**) en fonction des itérations – RSB=10dB – (gauche) $\widehat{M} = 3$ - (droite) $\widehat{M} = 4$

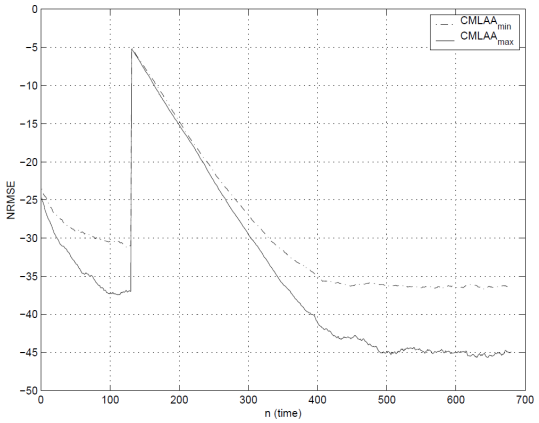


FIGURE 2.5 – Capacité de poursuite du CMLAA$_{min}$ et du CMLAA$_{max}$ (RSB=30dB, 20 réalisations).

séquences (*bursts*) de longueur suffisamment faible pour que l'on puisse considérer que le canal reste constant tout au long de la séquence. Il suffit alors d'introduire, au centre de la séquence, un paquet de symboles connus pour pouvoir estimer le canal puis les données envoyées. Ce mode de fonctionnement est remis en cause dès que la vitesse devient trop importante. Par exemple, pour un téléphone GSM se déplaçant à une vitesse de 300 Km/h (pour une fréquence porteuse de 1800MHz), le temps de cohérence du canal sera de 0.8ms. Cela implique des variations du canal non négligeables pendant la procédure d'égalisation du canal. Le projet de partenariat pour la $3^{ème}$ génération (3GPP) a donné des bornes sur la vitesse entre émetteur et récepteur au dessous de laquelle une certaine qualité de transmission doit être garantie [166] [6]. Toutefois, le standard ne requiert pas de performances spécifiques pour des vitesses supérieures à 250km/h pour des porteuses à 900MHz. Les récepteurs usuels basés sur l'algorithme de Viterbi ont des performances tout à fait convenables dans les situations prévues par le standard [41, 169] toutefois certaines applications nécessitent de travailler à des vitesses supérieures. L'objectif que nous nous étions fixés était de trouver un algorithme qui affiche de meilleures performances que l'algorithme de Viterbi dans les situations à très haute vitesse et des performances au moins similaires dans les autres situations.

6. Spécifications données en 2003, date à laquelle ont été menés ces travaux.

L'algorithme CMLAA$_\gamma$ a de nombreuses qualités qui en font un bon point de départ pour une utilisation dans le contexte GSM. C'est un algorithme adaptatif issu d'un critère de type Maximum de Vraisemblance connu pour être compétitif lorsque très peu de données dont disponibles. La formulation choisie permet l'introduction de connaissances *a priori* sur les symboles mais aussi sur le canal. Son inconvénient principal est l'existence de minima locaux. La convergence vers de tels minima est ici extrêmement improbable puisque nous disposons d'une séquence d'apprentissage qui fournit une initialisation. L'usage du CMLAA$_\gamma$ nécessite toutefois un minimum de diversité. Le paragraphe suivant explique comment obtenir une modélisation de type SIMO dans le contexte GSM.

Le système GSM utilise une modulation GMSK (Gaussian Minimum Shift Keying). Cette modulation est aussi utilisée pour GPRS (General Packet Radio Service) et pour EGPRS (Enhanced GPRS, modulation et schéma de codage de 1 à 4) [94]. C'est cette modulation qui va nous permettre de modéliser le système par un schéma de type SIMO avec une entrée et deux sorties. Les techniques habituelles d'obtention d'un modèle à plusieurs sorties ne sont pas envisageables (le coût d'une antenne supplémentaire était prohibitif et l'excès de bande insuffisant pour permettre le sur-échantillonnage) [166]. C'est la décomposition linéaire de Laurent [122] qui permet de voir le récepteur comme une structure linéaire de type SIMO. Ce modèle a été utilisé dans la littérature par divers auteurs [67, 117]. Le signal reçu s'écrit alors :

$$x(n) = \sum_k h(k)s(n-k) + b(n) \tag{2.34}$$

avec $h(k) = (h_0 * h_p * g)(k).j^{-k}$ où h_0 est un filtre provenant de la décomposition linéaire de la GMSK, h_p est le canal de propagation et g est le filtre de réception et où j^{-k} traduit une opération de rotation. Sachant que $h(k)$ est à valeurs complexes et que $\{s(n)\}$ est issu d'une modulation BPSK réelle, $x(n)$ peut être vu comme la sortie d'un système à 2 canaux, l'un correspondant à la partie réelle, l'autre à la partie imaginaire.

Nous avons dans un premier temps évalué les performances des algorithmes présentés plus haut dans le contexte GSM. Les conditions expérimentales sont les suivantes.

Conditions expérimentales 2.3 Nous considérons des paquets de données contenant 61 symboles inconnus suivis d'une séquence d'apprentissage de longueur 26 suivie de 61 symboles inconnus [166]. La séquence d'apprentissage fournit une initialisation aux algorithmes. ∎

La comparaison porte sur l'algorithme de Viterbi et sur les versions semi-aveugles du CMLBA$_{max}$ (algorithme bloc non partitionné), du CMLBA$_{max}$(12) [7] et du CMLAA (algorithme adaptatif). Nous avons constaté les choses suivantes. A haute vitesse (300km/h, 1800MHz) tous nos algorithmes ont de meilleures performances que l'algorithme de Viterbi. L'algorithme qui a les meilleures performances est le CMLBA partitionné. Nous répétons l'expérience pour un canal statique et nous constatons une importante dégradation lorsque l'on utilise l'algorithme adaptatif. Nous avons donc travaillé sur un nouvel algorithme adaptatif issu non pas du CMLBA mais du CMLBA partitionné.

Nous avons procédé à plusieurs modifications qui sont données sur la figure 2.6. Nous considérons une fenêtre w qui schématise l'étendue temporelle concernée par la minimisation. A un instant donné, les symboles $\mathbf{s}_u$ sont les symboles inconnus et constituent donc la variable d'optimisation. Les autres symboles sont considérés comme estimés (itération précédente). Le critère fait usage des deux interférences $I(\mathbf{s}_p;\mathbf{h})$ et $I(\mathbf{s}_f;\mathbf{h})$. Une fenêtre de pondération est utilisée avec facteur d'oubli λ comme montré sur la figure 2.6. L'interférence passée et future est ainsi pondérée de manière symétrique. La minimisation est ensuite itérée [8]. Nous nommons CML ce nouvel algorithme.

7. Un choix de 12 partitions conduit à des blocs de longueur 5 ce qui est exactement la longueur du canal.

8. On utilise ici un double système d'itérations, un des deux indice correspond au symbole courant au sein de la séquence et le deuxième compte le nombre de passages sur la séquence puisque l'on balaie le même paquet de manière répétée.

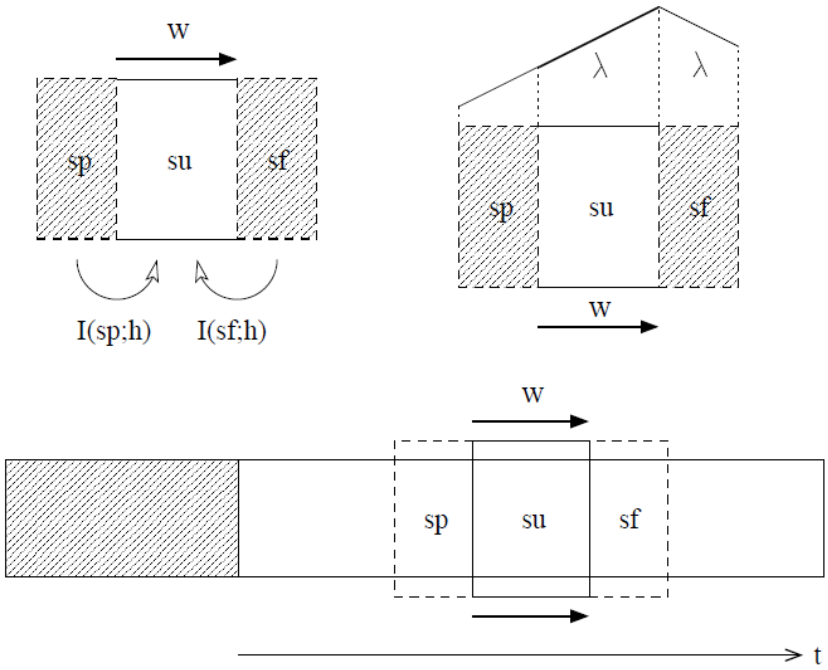
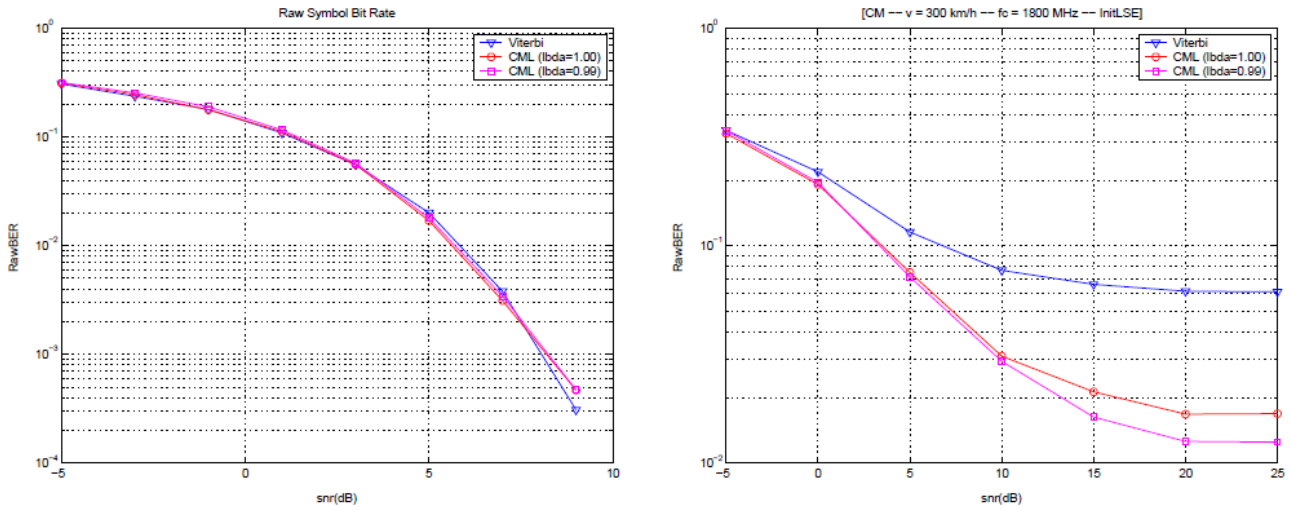


FIGURE 2.6 – Schéma de principe de l'algorithme proposé pour le GSM

Les résultats expérimentaux sont donnés sur la figure 2.7. On peut voir que le CML a des performances très similaires à celle de l'algorithme de Viterbi à vitesse nulle et bien meilleures à haute vitesse. On observe en particulier un gain de 5dB avec un facteur d'oubli $\lambda = 0.99$ par rapport à l'algorithme de Viterbi pour un RBER cible de $7.5.10^{-2}$. Le procédé proposé a fait l'objet d'un dépôt de brevet et d'une publication [78].



(a) Vitesse nulle (b) Haute vitesse

FIGURE 2.7 – Comparaison du CML et du Viterbi pour différentes vitesses.

2.1.3 Canal OFDM

Le travail présenté dans cette section porte à nouveau sur l'estimation canal/symbole avec canal variant dans le temps. Le système considéré n'est plus multi-canal mais multi-porteuse. Nous allons dans ce chapitre commencer par expliquer les caractéristiques du système considéré puis nos choix et notre contribution. Les travaux présentés ici sont relatifs à la thèse de Jean-Marcel Mamfoumbi et ont été menés entre 2002 et 2006. Ils ont conduit à trois communications dans des conférences internationales **[12, 157, 185]**

Nous considérons ici un système OFDM. Cette technique multi-porteuse présente de nombreux avantages dont la robustesse aux évanouissements par trajets multiples, un débit potentiellement élevé, une haute efficacité spectrale et une facilité d'implémentation. Pour toutes ces raisons, de nombreuses normes et standards (DAB, DVB, WIMAX, HIPERLAN/2, IEEE 802.11a,...) de communications en recommandent l'usage. Pour estimer correctement les données transmises, le récepteur OFDM doit posséder une estimation fiable du canal ce qui peut s'avérer délicat lorsque le canal est variant dans le temps. Usuellement, on utilise une séquence d'apprentissage qui permet d'obtenir une estimée du canal, on peut également utiliser toute l'information *a priori* à notre disposition pour affiner ou rafraîchir l'estimation du canal. Nous considérons ici un système de type WLAN (*Wireless Local Area Network*) qui est, par nature, un système dans lequel les utilisateurs peuvent se déplacer engendrant un canal de transmission non-statique. L'identification du canal peut être réalisée soit dans le domaine temporel [124, 151, 182] soit dans le domaine fréquentiel [70, 130, 131, 186]. Chaque alternative a été explorée donnant lieu a de nombreuses contributions dont un nombre important a été réalisé à la même époque ou postérieurement aux travaux présentés ici. L'avantage principal de l'estimation dans le domaine temporel est de réduire le nombre de coefficients à estimer. En effet, la longueur de la réponse impulsionnelle du canal est en général inférieure à celle du préfixe cyclique qui est lui-même de longueur inférieure à $\frac{P}{4}$ où P est le nombre de sous-porteuses. Le prix à payer est un estimation plus complexe et relativement coûteuse puisque l'on perd, dans le domaine temporel, la structure diagonale qui est l'avantage principal de l'estimation dans le domaine fréquentiel. Nous avons choisi ici de travailler dans le domaine fréquentiel et de proposer une estimation compatible avec une prise en compte du modèle d'évolution du canal qui est partiellement connu. Nous exploiterons également l'existence de symboles pilotes en utilisant une méthode semi-aveugle. Dans le cas du GSM, nous nous intéressions à une situation haute vitesse. Nous considérons ici un canal à variation lente ce qui conduit à des solutions algorithmiques différentes.

Système considéré et hypothèses

Nous considérons un système OFDM classique tel que les P_i porteuses centrales transportent de l'information. Le bloc de données est modulé par des porteuses orthogonales par l'intermédiaire de la transformée de Fourier inverse. Chaque composante est ensuite envoyée séquentiellement sur le canal après l'ajout d'un préfixe cyclique de longueur L. La longueur de l'intervalle de garde est supposée plus grande que la longueur du canal donc, dans le domaine fréquentiel, le canal peut être vu comme P_i canaux non sélectifs en parallèle. Les coefficients du canal sont supposés indépendants et distribués suivant une distribution de Rayleigh. Soit k un indice de sous-porteuse avec $0 \leq k \leq P_i - 1$, on a

$$y_k = h_k x_k + e_k \tag{2.35}$$

où y_k est le signal reçu sur la sous-porteuse k, x_k est un symbole inconnu à déterminer, e_k est un bruit additif gaussien de moyenne nulle et de variance σ_e^2 et h_k est le coefficients du canal. Nous considérons l'ensemble des hypothèses suivantes :

- (H_1) Le gain h_k est supposé constant durant K symboles OFDM. La valeur de K dépend de la vitesse du dispositif mobile. C'est un compromis à ajuster entre qualité de l'estimation et

prise en compte de la situation réelle. Soit H_n le coefficient du $n^{ième}$ paquet. Le vecteur d'observation $\mathbf{y} = (y_0, y_1, ..., y_{N-1})^T$ s'écrit :

$$\mathbf{y} = (H_0 x_0 \dots H_0 x_{K-1} \dots H_{q-1} x_{(q-1)K} \dots H_{q-1} x_{N-1})^T + \mathbf{e} \tag{2.36}$$

où $N = qK$ et $\mathbf{e} = (e_0, e_1, ..., e_{N-1})^T$. Avec des notations matricielles, on a

$$\mathbf{y} = \mathbf{XBH} + \mathbf{e} \tag{2.37}$$

où $\mathbf{H} = (H_0, H_1, ..., H_{q-1})^T$, $\mathbf{X} = Diag(x_0, x_1, ..., x_{N-1})$ contient les symboles d'information sur la diagonale et $\mathbf{B}$ est une matrice de taille $N \times q$ obtenue à partir du produit de Kronecker $I_q \otimes [1, ..., 1]^T$ où I_q est la matrice identité de taille q.

(H_2) Nous considérons un modèle lentement variant [9] dans le temps pour lequel les variations du canal sont modélisées à l'aide d'un modèle auto-régressif (AR). On considère ici un modèle d'ordre 1 mais les algorithmes proposés peuvent être étendus sans difficulté à des modèles d'ordre supérieur. Le modèle d'évolution de H_n est donné par

$$H_n = \alpha H_{n-1} + \varepsilon_n \tag{2.38}$$

où ε_n est un bruit additif gaussien de moyenne 0 et de variance σ_ε^2. Le coefficient α est supposé connu ou estimé. Dans la littérature, les canaux lentement variants sont très souvent modélisés par des modèles AR d'ordre 1 lorsque les canaux sont fortement corrélés [8, 103, 187]. Les modèles d'ordre faible sont alors en mesure de retranscrire l'essentiel de la dynamique du canal et conduisent à des algorithmes efficaces pour suivre les variations du canal [115].

(H_3) Les données sont supposées uniformément distribuées et à valeurs dans une constellation de taille M $\{s_1, ..., s_M\}$.

(H_4) Des symbole pilotes sont disponibles. Nous proposons donc une méthode semi-aveugle pour laquelle les symboles pilotes sont utilisés pour obtenir une estimée initiale du canal.

Algorithme d'estimation semi-aveugle

Notation 2.3. *Dans toute la section 2.1.3, la notation* $\mathbf{A}^*$ *désignera la matrice transposée de la matrice conjuguée d'une certaine matrice* $\mathbf{A}$.

L'algorithme d'estimation retenu est l'algorithme EM [64, 145], il est couramment utilisé dans ce genre d'application [10] [104, 130, 151, 206]. Il permet un couplage avec des algorithmes itératifs fournissant les probabilités *a posteriori* des symboles, il permet également l'introduction de connaissances *a priori* sur le canal ou sur les symboles. Les estimateurs au sens du maximum de vraisemblance sont optimaux asymptotiquement, nous l'avons mentionné dans la section 2.1.2. Précédemment, nous avions opté pour une version déterministe du maximum de vraisemblance pour limiter la complexité de la procédure d'estimation. Nous étions dans une situation à haute vitesse et nous considérions l'estimation de plusieurs canaux contenant chacun plusieurs coefficients. Ici, nous cherchons à estimer un canal mono-trajet (par porteuse) et nous nous plaçons dans une situation de faible à moyenne vitesse où le canal est supposé constant sur une certaine fenêtre temporelle. La complexité calculatoire est donc acceptable. Les approches développées dans [104, 206] n'introduisent pas, dans l'algorithme EM, les spécificités du système OFDM. En revanche, [130] utilise la connaissance des coefficients de canal des autres porteuses sous la forme

9. Avec cette hypothèse, nous pouvons considérer comme négligeable l'interférence entre porteuses.

10. Les références données dans cette partie correspondent à des contributions publiées au moment où nous avons débuté ce travail, de nombreuses contributions ont vu le jour depuis.

d'observations supplémentaires. Ce travail est complété par l'introduction d'un modèle AR d'ordre 1 (celui que nous avons décrit plus haut) qui permet de suivre les variations du canal [131]. Notre contribution se situe dans la continuité de ces travaux et portent sur la complexité calculatoire de l'algorithme. Nous montrons qu'il est possible d'obtenir une méthode d'estimation inspirée de l'EM à complexité linéaire et incluant des informations *a priori*.

Notre démarche est la suivante. Dans un premier temps, la connaissance du modèle d'évolution du canal n'est pas introduite dans la procédure d'estimation. Nous montrons alors que l'algorithme proposé a, pour chaque porteuse, une complexité linéaire, proportionnelle à NM. Nous introduisons ensuite l'information *a priori* dans le critère ce qui a pour effet indésirable une augmentation de la complexité de l'algorithme qui devient quadratique. Nous montrons enfin comment il est possible de modifier la procédure itérative afin d'obtenir à nouveau un algorithme à complexité linéaire. Ces trois étapes sont détaillées ci-dessous.

Le principe de l'algorithme EM est brièvement rappelé. L'EM est un algorithme itératif en deux étapes permettant de trouver le maximum de vraisemblance des paramètres d'un système en présence de données manquantes ou inobservables. La vraisemblance augmente à chaque étape garantissant une amélioration de la fiabilité de l'estimateur avec les itérations. La convergence vers le maximum global n'est pas garantie et peut dépendre de l'initialisation. L'algorithme présenté ici est obtenu directement à partir de l'algorithme EM appliqué au système décrit plus haut en utilisant l'hypothèse 1 (canal constant par morceaux) mais pas l'hypothèse 2 (modèle d'évolution du canal). La variable cachée est ici x_i, qui est considérée uniformément distribuée et à valeurs dans $\{s_0,...,s_{M-1}\}$ (hypothèse 3). Par conséquent l'ensemble des valeurs possibles de $X = Diag(x_0,...,x_{N-1})$ est donné par $\mathbf{S}_{\underline{m}} = Diag(s_{m_0},...,s_{m_{N-1}})$, $\underline{m} = [m_1,...,m_N]^T \in \{1,...,M\}^N$. Puisque le bruit additif est gaussien de moyenne nulle et de variance σ_e^2, la vraisemblance (complète) s'écrit :

$$P(\mathbf{y},\mathbf{S}_{\underline{m}} \mid \mathbf{H};\sigma_e^2) = P(\mathbf{S}_{\underline{m}})\frac{1}{\sigma_e^{2N}} exp(-\frac{1}{\sigma_e^2}\|\mathbf{y}-\mathbf{S}_{\underline{m}}\mathbf{B}\mathbf{H}\|^2) \tag{2.39}$$

On peut en déduire les équations de remise à jour suivantes [131] :

$$\mathbf{D}^{(i)}\mathbf{H}^{(i+1)} = \mathbf{V}^{(i)} \tag{2.40}$$

$$\sigma_e^{2(i+1)} = \frac{1}{N\Gamma^{(i)}}\sum_{\underline{m}} P(\mathbf{S}_{\underline{m}}|\mathbf{y},\mathbf{H}^{(i)};\sigma_e^{2(i)})\|\mathbf{y}-\mathbf{S}_{\underline{m}}\mathbf{B}\mathbf{H}^{(i)}\|^2 \tag{2.41}$$

où (i) indique un numéro de l'itération, $\mathbf{D}^{(i)}$ est une matrice, $\mathbf{V}^{(i)}$ un vecteur et $\Gamma^{(i)}$ un scalaire ayant pour expression :

$$\mathbf{D}^{(i)} = \sum_{\underline{m}} P(\mathbf{S}_{\underline{m}}|\mathbf{y};\mathbf{H}^{(i)},\sigma_e^{2(i)})\mathbf{B}^*\mathbf{S}_{\underline{m}}^*\mathbf{S}_{\underline{m}}\mathbf{B} \tag{2.42}$$

$$\mathbf{V}^{(i)} = \sum_{\underline{m}} P(\mathbf{S}_{\underline{m}}|\mathbf{y};\mathbf{H}^{(i)},\sigma_e^{2(i)})\mathbf{B}^*\mathbf{S}_{\underline{m}}^*\mathbf{y} \tag{2.43}$$

$$\Gamma^{(i)} = \sum_{\underline{m}} P(\mathbf{S}_{\underline{m}}|\mathbf{y};\mathbf{H}^{(i)},\sigma_e^{2(i)}) \tag{2.44}$$

Ecrit de cette façon, la somme sur le multi-index $\underline{m}$ dans les expressions de $\mathbf{D}^{(i)}$ et $\mathbf{V}^{(i)}$ conduit à une complexité exponentielle. Mais, nous pouvons ici tirer profit des deux propriétés suivantes liées à la définition de la matrice $\mathbf{B}$.

Propriété 2.1.2 Soit $\mathbf{B}$ la matrice de taille $N \times q$ avec $N = qK$ et définie selon $\mathbf{B} = I_q \otimes [1,...,1]^T$ où I_q est la matrice identité de taille q et $\otimes$ est le produit de Kronecker. Pour tout vecteur $(z_0,\ldots,z_{N-1})^T$ à valeur réelles ou complexes, on a

$$\mathbf{B}^* Diag(z_0,...,z_{N-1})\mathbf{B} = Diag(z_0+...+z_{K-1},...,z_{(q-1)K}+...+z_{N-1}) \tag{2.45}$$

$$\mathbf{B}^*[z_0,...,z_{N-1}]^T = [z_0+...+z_{K-1},...,z_{(q-1)K}+...+z_{N-1}]^T \tag{2.46}$$

En utilisant ces deux éléments et l'indépendance des symboles, la remise à jour des coefficients du canal pour $j=0,...,q-1$ est donnée par :

$$H_j^{(i+1)} = \frac{\sum_{l=jK}^{(j+1)K-1} y_l \frac{\sum_{m=1}^{M} s_m^* P(s_m)\exp(-\frac{1}{\sigma_e^{2(i)}}|y_l - H_j^{(i)}s_m|^2)}{\sum_{m=1}^{M} P(s_m)\exp(-\frac{1}{\sigma_e^{2(i)}}|y_l - H_j^{(i)}s_m|^2)}}{\sum_{l=jK}^{(j+1)K-1} \frac{\sum_{m=1}^{M} |s_m|^2 P(s_m)\exp(-\frac{1}{\sigma_e^{2(i)}}|y_l - H_j^{(i)}s_m|^2)}{\sum_{m=1}^{M} P(s_m)\exp(-\frac{1}{\sigma_e^{2(i)}}|y_l - H_j^{(i)}s_m|^2)}} \tag{2.47}$$

et celle de la variance du bruit par

$$\sigma_e^{2(i+1)} = \frac{1}{N}\sum_{j=0}^{q-1}\sum_{l=jK}^{(j+1)K-1}\sum_{m=1}^{M} \frac{P(s_m)\exp(-\frac{1}{\sigma_e^{2(i)}}|y_l - H_j^{(i)}s_m|^2)}{\sum_{m'=1}^{M} P(s_{m'})\exp(-\frac{1}{\sigma_e^{2(i)}}|y_l - H_j^{(i)}s_{m'}|^2)} |y_l - H_j^{(i+1)}s_m|^2 \tag{2.48}$$

On pourra consulter [185] pour les détails des calculs. Les équations de remise à jour (2.47) et (2.48) ont une complexité arithmétique proportionnelle à $qKM = NM$ c'est à dire linéaire avec le nombre de variables à estimer. C'est un résultat intéressant qui conforte le choix d'un algorithme EM dans le contexte étudié. Nous noterons dans la suite EM-Bloc, l'algorithme (2.47-2.48). Le modèle d'évolution du canal est maintenant pris en compte dans la procédure d'estimation. La relation (2.38) s'écrit sous forme matricielle selon :

$$\mathbf{MH} = \boldsymbol{\varepsilon} + \mathbf{b} \tag{2.49}$$

où $\mathbf{b}$ est un vecteur de longueur q ayant pour première composante une valeur initiale du coefficient du canal calculée à partir du paquet de données précédent [11]. La fonction de vraisemblance avec données complètes reste inchangée (cf. eq (2.39)). A partir de (2.49), on obtient la probabilité $P(\mathbf{H})$ ci-après :

$$P(\mathbf{H}) \propto \frac{1}{\sigma_\varepsilon^{2q}} \exp\left(-\frac{1}{\sigma_\varepsilon^2}(\mathbf{H}-\mathbf{M}^{-1}\mathbf{b})^*\mathbf{C}^{-1}(\mathbf{H}-\mathbf{M}^{-1}\mathbf{b})\right) \tag{2.50}$$

où $\sigma_\varepsilon^2\mathbf{C}$ est la matrice de covariance de $\mathbf{H}$ et $\mathbf{C}^{-1} = \mathbf{M}^*\mathbf{M}$. En oubliant les termes constants vis à vis des paramètres du problème, la fonction auxiliaire se réécrit sous la forme suivante :

$$Q(\mathbf{H},\sigma_e^2,\mathbf{H}^{(i)},\sigma_e^{2(i)}) = \sum_{\underline{m}} P(\mathbf{S}_{\underline{m}}|\mathbf{y};\mathbf{H}^{(i)},\sigma_e^{2(i)})\left(\log P(\mathbf{y}|\mathbf{S}_{\underline{m}},\mathbf{H};\sigma_e^2) + \log P(\mathbf{H})\right)$$

11. La matrice $\mathbf{M}$ de taille $q\times q$ est donnée ci-dessous, son inverse que nous utiliserons dans la suite du document est également donnée :

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -\alpha & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\alpha & 1 \end{pmatrix} \text{ and } \mathbf{M}^{-1} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \alpha & 1 & \ddots & \ddots & \vdots \\ \alpha^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \alpha^{q-1} & \cdots & \alpha^2 & \alpha & 1 \end{pmatrix}$$

donnant lieu à une estimée au sens du maximum *a posteriori*. L'ajout du logarithme de la densité *a priori* améliore en général la concavité de la fonction d'objectif [133]. Après calcul puis annulation du gradient de $Q(\mathbf{H},\sigma_e^2,\mathbf{H}^{(i)},\sigma_e^{2(i)})$ par rapport à $\mathbf{H}^*$, nous obtenons :

$$\left(\frac{1}{\sigma_e^2}\mathbf{D}^{(i)}+\frac{1}{\sigma_\varepsilon^2}\mathbf{C}^{-1}\Gamma^{(i)}\right)\mathbf{H}^{(i+1)}=\frac{1}{\sigma_e^2}\mathbf{V}^{(i)}+\frac{1}{\sigma_\varepsilon^2}\mathbf{M}^*\mathbf{b}\Gamma^{(i)} \tag{2.51}$$

où $\mathbf{D}^{(i)}$, $\mathbf{V}^{(i)}$ et $\Gamma^{(i)}$ ont déjà été définis dans (2.42)-(2.44). L'estimée du canal est obtenue en résolvant (2.51) à l'aide d'une méthode de Gauss. Comme $\mathbf{C}^{-1}$ n'est pas une matrice diagonale, la résolution de (2.51) a une complexité quadratique. Nous verrons dans la suite que les performances obtenues par ce dernier algorithme que nous noterons EM-MAP sont très supérieures à celles obtenues avec l'algorithme EM-Bloc. Nous montrerons aussi que nous pouvons conserver le bénéfice de cette amélioration sans augmenter le coût de calcul. Nous devons pour cela modifier les équations de mise à jour. Nous utilisons une technique connue sous le nom de *One Step Late* (OSL) et proposée par Green dans [87]. L'idée est simple ; elle consiste à évaluer la dérivée de l'*a priori* $\log(P(\mathbf{H}))$ non plus au point courant $\mathbf{H}^{(i+1)}$ mais au point $\mathbf{H}^{(i)}$ obtenu à l'itération précédente. En appliquant, cette technique à (2.51), nous obtenons la nouvelle équation de remise à jour :

$$\frac{1}{\sigma_e^2}\mathbf{D}^{(i)}\mathbf{H}^{(i+1)}=\frac{1}{\sigma_e^2}\mathbf{V}^{(i)}+\frac{1}{\sigma_\varepsilon^2}\Gamma^{(i)}\mathbf{M}^*\mathbf{b}-\frac{1}{\sigma_\varepsilon^2}\Gamma^{(i)}\mathbf{C}^{-1}\mathbf{H}^{(i)} \tag{2.52}$$

Comme pour l'EM-Bloc (2.40), la matrice à inverser pour résoudre (2.52) est diagonale. La complexité arithmétique est donc à nouveau linéaire. Clairement, la technique OSL ne change pas les points fixes de l'algorithme. En revanche, dans le cas général une procédure OSL ne garantit pas une augmentation de la log-vraisemblance pénalisée au fil des itérations et il n'existe pas de preuve générale de convergence. Nous devrons donc prouver la convergence de la procédure proposée. Pour que la complexité arithmétique occasionnée par le calcul de $\mathbf{H}^{(i+1)}$ soit linéaire, il faut que la matrice qui multiplie ce vecteur (et qui doit être inversée) soit diagonale. Nous introduisons un degré de liberté supplémentaire sous la forme d'un paramètre β qui ajuste la quantité d'éléments diagonaux pouvant être réaffectés au membre de gauche de (2.52). Plus précisément, nous proposons de modifier (2.52) selon :

$$\left(\frac{1}{\sigma_e^2}\mathbf{D}^{(i)}+\frac{\beta}{\sigma_\varepsilon^2}\Gamma^{(i)}\mathbf{I}\right)\mathbf{H}^{(i+1)}=\frac{1}{\sigma_e^2}\mathbf{V}^{(i)}+\frac{1}{\sigma_\varepsilon^2}\Gamma^{(i)}M^*\mathbf{b}-\frac{1}{\sigma_\varepsilon^2}\Gamma^{(i)}\mathbf{C}^{-1}\mathbf{H}^{(i)}+\frac{\beta}{\sigma_\varepsilon^2}\Gamma^{(i)}\mathbf{H}^{(i)} \tag{2.53}$$

On retrouve les deux mêmes propriétés que pour (2.52) : une complexité linéaire et des points fixes identiques à ceux de l'EM-MAP. Nous nommons ce nouvel algorithme EM-OSL. Le paramètre β va s'avérer essentiel pour prouver la convergence et pour ajuster la vitesse de convergence de manière optimale. Pour cela nous allons nous intéresser à l'algorithme du point proximal que nous noterons PPA dans la suite pour *Proximal Point Algorithm*. Cette classe d'algorithmes a été introduite par Rockafellar [173] et Martinet [128] en s'inspirant des travaux de Minty [142] et Moreau [147]. Dans ce document, nous définissons un algorithme de type PPA de la même manière que [50].

Définition 2.1.1 Un algorithme (généralisé) du point proximal est défini par le processus itératif

$$\Theta^{(i+1)}=\arg\max_{\Theta}\{f(\Theta)-\gamma_i d(\Theta,\Theta^{(i)})\} \tag{2.54}$$

où γ_i est une séquence de nombres positifs et où $d(\Theta,\Theta^{(i)})$ est un terme de pénalité tel que, pour

tout couple de paramètres $(\Theta, \Theta^{(i)})$:

$$d(\Theta,\Theta^{(i)}) \geq 0 \quad (2.55)$$
$$d(\Theta,\Theta^{(i)}) = 0 \text{ si et seulement si } \Theta = \Theta^{(i)} \quad (2.56)$$

Si ces deux conditions sont remplies alors $\{f(\Theta^{(i)})\}$ est une séquence non-décroissante.

Le lien entre l'algorithme EM et la formulation PPA a été établi dans [72] et plus tard dans [50] conduisant à une preuve alternative de la convergence de l'EM. L'algorithme EM-MAP (2.51) est aussi un PPA tel que $f(\Theta) = \log\ p(\mathbf{y}|\Theta) + \log\ p(\Theta)$, $\gamma_i = 1$ et

$$d(\Theta,\Theta^{(i)}) = E\left[\log \frac{p(\mathbf{x}|\mathbf{y};\Theta^{(i)})}{p(\mathbf{x}|\mathbf{y};\Theta)} |\mathbf{y};\Theta^{(i)}\right] \quad (2.57)$$

Cette distance est une divergence de Kullback ; elle vérifie donc les conditions (2.55-2.56), $f(\Theta)$ est la log-probabilité *a posteriori* donc après chaque étape de l'EM-MAP on obtient une suite d'estimations telles que les probabilités *a posteriori* associées forment une suite non-décroissante. Ce résultat n'est pas nouveau mais nous avons ici une méthode qui va nous être utile pour prouver la convergence de l'EM-OSL. On montre le résultat suivant [157].

Résultat 2.1.2 L'algorithme EM-OSL décrit par l'équation (2.53) peut s'écrire sous la forme :

$$\Theta^{(i+1)} = \arg\max_{\Theta}(f(\Theta) - \gamma_i d'(\Theta,\Theta^{(i)})) \quad (2.58)$$

où la fonction $f(\Theta)$ est la log-probabilité *a posteriori*, $\gamma_i = 1$ et $d'(\Theta,\Theta^{(i)})$ est donné par

$$d'(\Theta,\Theta^{(i)}) = d(\Theta,\Theta^{(i)}) + \frac{1}{\sigma_\varepsilon^2}(\mathbf{H}-\mathbf{H}^{(i)})^*(\beta\mathbf{I}-\mathbf{C}^{-1})(\mathbf{H}-\mathbf{H}^{(i)}) \quad (2.59)$$

où $d(\Theta,\Theta^{(i)})$ est la distance donnée dans (2.57) pour l'EM-MAP.

Une condition suffisante pour prouver la convergence de l'EM-OSL est de choisir β tel que $\beta\mathbf{I}-\mathbf{C}^{-1}$ soit une matrice définie positive ainsi $d'(\Theta,\Theta^{(i)})$ satisfera les conditions (2.55-2.56). A partir de la distributions des valeurs propres de $\mathbf{C}^{-1}$, on obtient la condition suffisante

$$\beta \geq 1 + \alpha^2 + 2\alpha\cos(\frac{\pi}{q+1}) \quad (2.60)$$

Cette borne est obtenue en considérant séparément les deux termes présents dans $d'(\Theta,\Theta^{(i)})$ et en exigeant de chacun d'eux qu'il respecte les conditions (2.55-2.56). Nous pouvons obtenir une borne mieux ajustée en travaillant sur la somme des deux termes. Plus précisément, nous travaillons sur le rayon spectral de la matrice gouvernant la convergence de l'algorithme. Classiquement [87], [134], [96], nous nous intéressons à la matrice

$$R_\beta = -[\nabla^{20}Q_{OSL}(\Theta,\Theta)]^{-1}\nabla^{11}Q_{OSL}(\Theta,\Theta) \quad (2.61)$$

où Q_{OSL} est la fonction auxiliaire de la première étape (étape E) de l'algorithme EM-OSL et où ∇^{ij} est un opérateur de dérivée partielle tel que $\nabla^{ij}F(x,y) = \frac{\partial^{i+j}F(x,y)}{\partial x^i \partial y^j}$. Les détails de calcul ainsi que les démonstrations sont données dans [12]. Nous obtenons une nouvelle borne sur β qui optimise la vitesse de convergence tout en garantissant la convergence de l'algorithme. Cette borne est donnée par

$$\beta_{min} = \frac{1 + \alpha^2 + 2\alpha\cos(\frac{\pi}{q+1})}{2} \quad (2.62)$$

Simulations

Nous donnons ici quelques résultats numériques obtenus pour le contexte de simulation donné ci-après.

Conditions expérimentales 2.4 Nous considérons ici le standard HIPERLAN 2 [53] pour lequel le préfixe cyclique est de longueur 16, le nombre de porteuses est égal à 64 dont 48 sont actives. Un code convolutif (171/173) de taux $R = 1/2$, de longueur de contrainte $l = 7$ est utilisé, les bits codés sont entrelacés et suivis d'un *mapping* vers une constellation 16-QAM. Un canal de type BRAN C est considéré [52]. ∎

Les résultats sont présentés pour un canal statique et pour des canaux variant dans le temps. Les vitesses sont choisies en accord avec les spécifications données par la norme. elles sont supposées connues au niveau du récepteur ce qui permet d'évaluer la valeur du coefficient de corrélation $\widetilde{\alpha}$. Dans les simulation, nous utilisons $\alpha = \widetilde{\alpha}^K$ puisque le canal est supposé constant par morceaux de K symboles. Chaque séquence possède deux symboles d'apprentissage suivis de 100 symboles OFDM. Les probabilités des bits individuels sont obtenues après deux itérations du processus de turbo-démodulation.

Nous comparons tout d'abord les trois algorithmes proposés entre eux pour une vitesse de 3m/s puis de 10m/s. Les tailles des blocs sont optimisées pour chacune des vitesses, par exemple pour une vitesse de 10m/s, le meilleur choix est $K = 10$. Les résultats sont donnés sur la figure 2.8. On

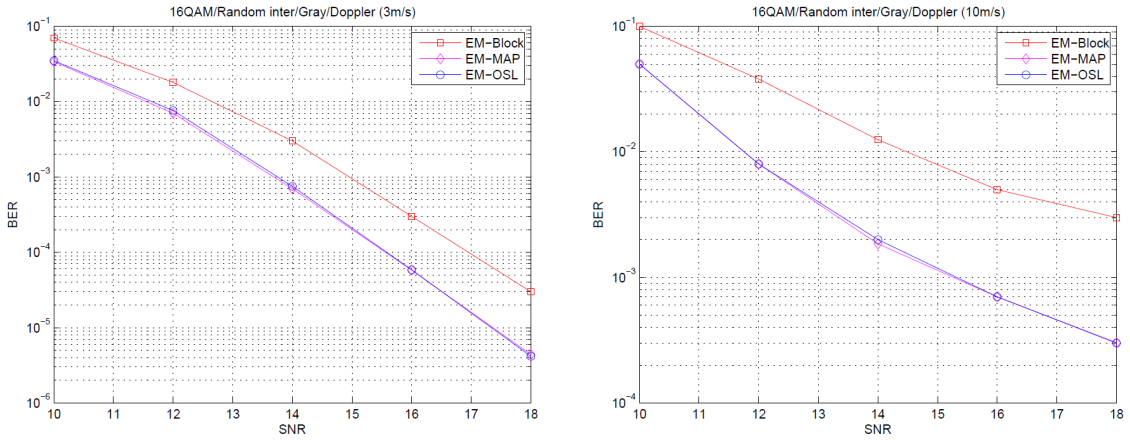


FIGURE 2.8 – Comparaison d'algorithmes en terme de BER pour une vitesse de 3m/s (gauche) puis de 10m/s (droite).

constate l'impact en terme de performances de l'introduction de l'*a priori* dans le critère. Sans surprise, l'EM-MAP et l'EM-OSL présentent des performances similaires avec l'avantage d'une plus faible complexité pour l'EM-OSL. La valeur choisie pour le degré de liberté β a un impact sur la vitesse de convergence. Nous avons comparé le gain en nombre d'itérations obtenu en utilisant pour β la valeur donnée en (2.62) par rapport à la valeur de β en (2.60). Nous obtenons, pour $\alpha = 0.997$, un gain en nombre d'itérations compris entre 21% et 36% pour des SNR allant de 10 à 18dB.

Nous comparons maintenant l'EM-OSL aux algorithmes suivants :

— EM-AR [131] : utilise également un modèle AR d'ordre 1 mais ne tient compte que d'une seule observation supplémentaire correspondant au coefficient du canal du symbole OFDM précédent sur la même porteuse.

— EM-OFDM [130] : utilise les corrélations fréquentielles et considère comme observations supplémentaires les coefficients du canal sur les autres porteuses.

— EM-OFDM(p) : identique à l'EM-OFDM mais les variances sont estimées dans l'EM-OFDM(p) alors qu'elles sont supposées connues dans l'EM-OFDM.
— LMS [92].

Les résultats sont donnés sur la figure 2.9 pour un canal statique et pour une vitesse de 3m/s. La courbe étiquetée *None* donne le résultat obtenu en utilisant la séquence d'apprentissage pour estimer le canal sans suivi des variations entre les séquences. Nous observons pour un taux binaire de 10^{-3}

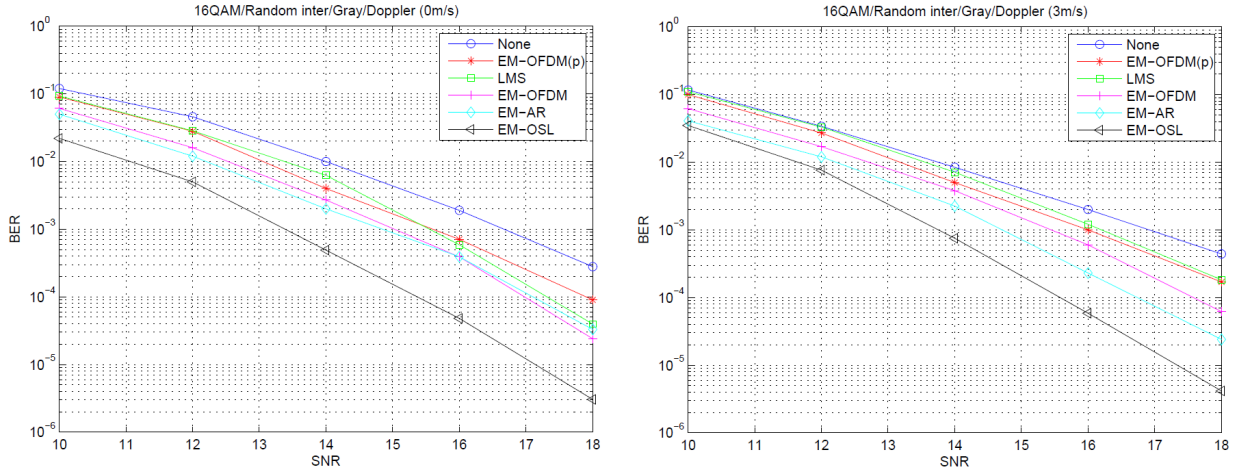


FIGURE 2.9 – Comparaison d'algorithmes en terme de BER pour une vitesse de 0m/s (gauche) puis de 3m/s (droite).

un gain de 3dB par rapport au schéma classique (courbe *None*) et un gain de 1 à 2dB par rapport à l'EM-AR.

2.1.4 Conclusion

J'ai présenté dans cette première partie une sélection de travaux sur les méthodes et techniques d'estimation adaptative de canaux de transmission non-statique. L'essentiel de ces travaux a été mené il y a plusieurs années. Le domaine des télécommunications est à évolution relativement rapide, j'ai donc replacé les travaux présentés dans le contexte de l'époque où ils ont été entrepris. Nous pouvons maintenant essayer de replacer les contributions dans le contexte d'aujourd'hui.

Les travaux présentés dans la sous-section 2.1.1 sont une analyse asymptotique d'une classe d'algorithmes adaptatifs à faible complexité. Notre contribution tient à la fois dans la méthodologie employée et dans les résultats obtenus pour ces algorithmes. Ces travaux s'inscrivent dans un ensemble de contributions plus vastes par le nombre d'algorithmes étudiés et aussi par les indices de performances considérés. Nous avons essentiellement travaillé sur la vitesse de convergence et l'écart à l'orthonormalité. D'autres auteurs se sont intéressés au problèmes d'instabilité liés aux erreurs d'arrondi en précision finie, aux stratégies à pas décroissant ou encore aux capacité de poursuite en environnement non-stationnaire. On pourra se référer à [9] pour une vision complète sur ces algorithmes adaptatifs et leurs performances.

Les travaux décrits dans la sous-section 2.1.2 ont été développés entre 1996 et 2006 époque où de nombreuses contributions ont vu le jour pour proposer des solutions aveugles et semi-aveugles pour la téléphonie mobile. La contribution principale est la proposition d'un algorithme du maximum vraisemblance adaptatif et dont la vitesse de convergence a été travaillée et améliorée en partitionnant efficacement la matrice de canal. Grâce à une collaboration avec la société Wavecom nous avons eu la possibilité d'évaluer l'applicabilité de la solution proposée. Contrairement aux travaux menés dans la section 2.1.1, nous n'avons pas proposé d'analyse analytique des performances

qui sont établies par simulation ce qui est aussi le cas pour la plupart des contributions alternatives proposées par d'autres auteurs à cette époque là. Une étude analytique a été faite ultérieurement, en 2013, par Carvalho *et al* dans [59] encouragés par le nouveau regain d'intérêt pour l'estimation aveugle de canal de transmission liés à l'émergence des recherches en radio cognitive. D'un point de vue applicatif, les techniques développées peuvent s'avérer utile pour de l'estimation aveugle de canal de transmission dans des réseaux de capteurs quelconques et pas nécessairement dédiés à la téléphonie mobile.

Enfin, la contribution principale de la sous-section 2.1.3 est essentiellement liée à l'EM-OSL et au travail que nous avons mené sur la complexité arithmétique ainsi que sur la vitesse de convergence. Dans des travaux non présentés ici, nous avons également utilisé avec succès les techniques du point proximal pour proposer des versions accélérées de l'algorithme de Blahut-Arimoto **[154],[155]**. Vis à vis d'une application en OFDM, nous avons pris en compte une partie de l'information disponible mais nous aurions pu faire mieux en intégrant d'autre caractéristiques. Au lieu de travailler porteuse par porteuse, nous aurions pu considérer également les porteuses adjacentes et intégrer la connaissance des corrélations fréquentielles. Nous avons pu montrer que les corrélations temporelles sont plus fortes que les corrélations fréquentielles c'est pour cela que nous avons considéré en premier lieu les corrélations temporelles toutefois l'incorporation d'informations *a priori* sur les corrélations fréquentielles auraient probablement conduit à des performances encore améliorées. De la même manière, nous avons supposé que la réponse impulsionnelle du canal est de longueur L, nous aurions pu exploiter le fait que la réponse en fréquence du canal vit dans un sous-espace de dimension L. Faute de temps, ces pistes initialement envisagées dans le sujet de thèse n'ont pu être explorées.

BILAN (THÈSE/PUBLICATION)

- 1 thèse soutenue (J-M. Mamfoumbi),
- 3 publication dans des revues internationales avec comité de lecture (*IEEE Trans. on Signal Processing* [R.7], [R.9] et [R.10]),
- 1 brevet,
- 8 publications dans les actes de conférences internationales avec comité de lecture,
- 2 publications dans les actes de conférences nationales avec comité de lecture.

2.2 Correction de bruit impulsionnel

Cette section rassemble une partie des travaux initiés lors de la thèse de F. Abdelkefi [1], soutenue en 2002. Nous nous sommes intéressés à la correction d'erreurs impulsionnelles dans les systèmes multi-porteuses. Le bruit impulsionnel est une perturbation additive dont la présence affecte la qualité des transmissions par son caractère purement aléatoire. Le système de correction que nous avons proposé, repose sur une analogie entre le modulateur OFDM (*Orthogonal Frequency-Division Multiplexing*) et un codeur de type Reed-Solomon dans le corps des complexes. La définition de codes correcteurs d'erreurs sur le corps des réels ou des complexes est présente dans les travaux de Marshall [126, 127] qui semble avoir été le premier à introduire les codes réels ou complexes basés sur la transformée de Fourier discrète. Il a prouvé que certaines propriétés des codes définis dans des corps de Galois sont encore valables et sont directement applicables dans le corps des complexes et des réels. Dans la même optique, Blahut a étendu la notion de code BCH calculé à partir d'une transformée de Fourier sur un corps fini à un corps infini [36, 37, 38].

L'utilisation de techniques de codage pour la correction de bruit impulsionnel a été considérée par plusieurs auteurs et pour des domaines d'application divers et variés [118, 129, 170, 195, 196]. Ces contributions considèrent que les séquences émises contiennent un nombre suffisant de

zéros consécutifs pouvant jouer le rôle de syndromes. Cette hypothèse est trop restrictive pour l'application à laquelle nous nous intéressons ici puisque les zéros en question sont situés sur une partie du spectre qui est atténuée par les filtres de mise en forme. En revanche, des symboles connus (les pilotes) sont présents à l'intérieur de chaque séquence dans le but de faciliter l'étape de synchronisation ou le processus d'égalisation. Une des originalités de notre travail est d'avoir montré que l'utilisation des symboles pilotes en tant que syndromes du codeur était possible.

La position des syndromes est directement liée à la capacité de correction. La borne BCH ainsi que la borne BCH généralisée [90, 174] donnent des conditions sur la position et sur le nombre de syndromes qui garantissent une certaine capacité de correction. Malheureusement, la position des symboles pilotes dans la séquence OFDM ne répond pas forcément aux critères stricts requis pour chacune de ces bornes. Nous avons donc cherché à établir une condition plus générale garantissant une certaine capacité de correction. Nous avons obtenu une condition nécessaire qui est aussi suffisante pour des capacités de correction égales à deux ou à trois.

La qualité de notre correcteur d'erreurs impulsionnelles peut être nettement améliorée en lui adjoignant un test de contrôle a posteriori. Pour tester les dysfonctionnements éventuels de notre algorithme, nous avons recours aux test d'hypothèses. Cette technique a aussi été exploitée par Redinbo dans [170]. Nous proposons une organisation des tests en cascade qui permet de calculer théoriquement les probabilités *a priori* pour chaque test considéré et de déduire les seuils à appliquer de manière plus fiable que dans [170] où ils sont fixés de manière empirique.

Nous avons ensuite évalué le gain apporté par l'algorithme de correction dans les deux situations suivantes : *(1)* correction d'erreurs impulsionnelles dans le contexte d'HIPERLAN 2, *(2)* réduction du niveau de PAPR (Peak to Average Power Ratio) toujours dans le contexte d'HIPERLAN 2. Le signal OFDM a en effet la particularité d'avoir une dynamique importante et est de plus soumis à la non-linéarité de divers dispositifs ce qui peut entraîner un écrêtement. Une solution intéressante évoquée dans [24] consiste à assimiler l'effet de l'écrêtement à un bruit de type impulsionnel. L'algorithme proposé peut être utilisé pour des écrêtements se produisant à la sortie du modulateur ou à l'entrée du démodulateur.

Nous organisons la présentation de ces travaux de la manière suivante. L'analogie entre le modulateur OFDM à base de transformée de Fourier et un codeur de type Reed-Solomon est d'abord présentée. Seront données ensuite les grandes lignes de l'algorithme de Peterson-Gorenstein-Zierler qui nous a servi de point de départ et qu'il me semble important de présenter pour une bonne compréhension du travail réalisé. Nos contributions sont ensuite présentées en mettant l'accent sur les modifications apportées à l'étape de décodage pour prendre en compte la position (quelconque) des pilotes ainsi que la présence du bruit de fond. Nous présenterons également les résultats obtenus concernant le lien entre la position des syndromes et la capacité de correction associée. En revanche, la réorganisation de l'algorithme en utilisant des tests d'hypothèse ne sera pas exposée ici. Pour plus de détails on pourra se reporter à [2, 3]. Quelques résultats numériques sont donnés sur la correction de bruit impulsionnel et la réduction du niveau de PAPR pour HIPERLAN 2.

2.2.1 OFDM et codes de Reed-Solomon

Le système OFDM

Nous considérons ici un système OFDM discret comme dans la partie 2.1.3. Les notations ne sont pas tout à fait les mêmes que dans la partie précédente. Pour plus de clarté, nous représentons le modèle discret du système OFDM sur la figure 2.10 avec les notations associées. Dans le modèle discret que nous considérons, le modulateur OFDM est décrit par les deux équations :

$$\mathbf{C} = (C_0, \ldots, C_{N-1}, 0, \ldots, 0)^T \tag{2.63}$$

$$\mathbf{c} = \mathbf{T}\mathbf{F}_M^{-1}\mathbf{C} \tag{2.64}$$

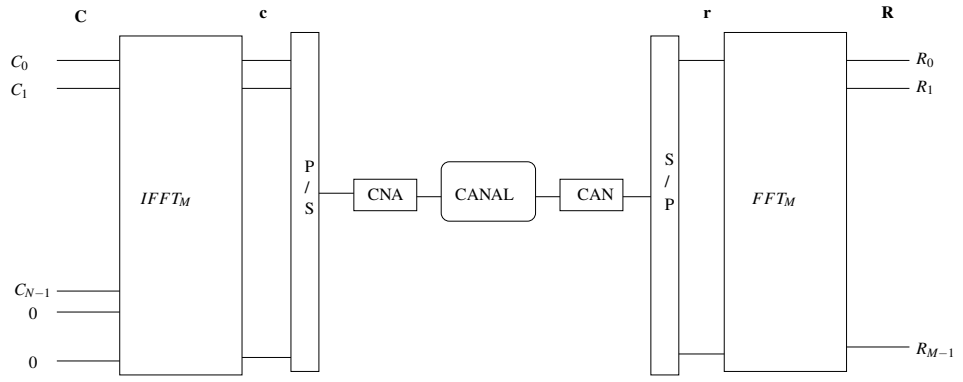


FIGURE 2.10 – Modèle discret d'un système OFDM

où $\mathbf{TF}_M^{-1}$ est la transformée de Fourier discrète inverse de taille M et où $\mathbf{C}$ est une séquence de longueur M contenant N symboles d'information et $M-N$ zéros. Au niveau du récepteur, après conversion analogique/numérique et après démodulation (FFT), les symboles reçus s'écrivent

$$R_k = C_k + N_k \quad 0 \leq k \leq N-1 \tag{2.65}$$

où N_k est la transformée de Fourier d'ordre M de la séquence de bruit $\{n_k\}$. Les symboles R_k pour $N \leq k \leq M-1$ sont les syndromes du code. Nous expliciterons cela dans la suite du document. Nous présentons maintenant les hypothèses sur le bruit de canal.

Le modèle de canal

Nous considérons un canal sans mémoire perturbé par un bruit gaussien et un bruit impulsionnel. Plusieurs modèles empiriques ont été proposés dans la littérature [34, 83, 139, 140]. Ces caractéristiques ont été formalisées par Gosh dans [83]. Dans tout les cas, le bruit impulsionnel a une variance très grande par rapport à celle du bruit de fond et une durée très faible. C'est ce modèle de canal que nous adoptons et qui s'écrit :

$$r_k \quad = \quad c_k + b_k + i_k \tag{2.66}$$

où $\mathbf{c}$ est la séquence émise perturbée par un bruit de fond additif $\mathbf{b}$, blanc, de moyenne nulle et de variance σ_b^2 et également par un bruit $\mathbf{i}$ de type Bernoulli-Gaussien.

Définition 2.2.1 Le bruit impulsionnel $\mathbf{i}$, Bernoulli-Gaussien, s'écrit :

$$i_k \quad = \quad \ell_k \, g_k \quad \forall k \in \{0,1,...,M-1\} \tag{2.67}$$

où ℓ est une séquence de Bernoulli, telle que : $prob(\ell_k = 1) = p$ et $prob(\ell_k = 0) = 1-p$ et $\mathbf{g}$ est une séquence de bruit gaussien, blanc, de moyenne nulle et de variance σ_i^2.

On considérera dans toute la suite que $\sigma_b^2 \ll \sigma_i^2$ et que la séquence de bruit impulsionnel et la séquence de bruit de fond sont indépendantes. Notons $\mathbf{n}$ le bruit total introduit par le canal, on a

$$n_k \quad = \quad b_k + i_k \tag{2.68}$$

La densité de probabilité de n_k est :

$$p_n \quad = \quad (1-p)\, G(n,0,\sigma_b^2) \; + \; p\, G(n,0,\sigma_b^2+\sigma_i^2) \tag{2.69}$$

où $G(n_k, m_x, \sigma_x^2)$ représente la densité de probabilité d'un bruit gaussien de moyenne m_x et de variance σ_x^2. Nous définissons maintenant les codes de Reed-Solomon qui font partie de la catégorie des codes en bloc et nous montrons l'analogie avec le système OFDM.

OFDM et codes de Reed-Solomon

Un code en bloc (M,N) est une association entre des mots d'information de longueur N et des mots de code de longueur M tel que $M > N$.

Définition 2.2.2 La distance minimale d d'un code est définie selon

$$d = \min_{i \neq j} d_H(c_i, c_j) \tag{2.70}$$

où c_i et c_j sont deux mots de code et $d_H(.,.)$ est la distance de Hamming. Cette distance est liée à la capacité de correction selon la relation

$$t = \lfloor \frac{d-1}{2} \rfloor \tag{2.71}$$

où t est le nombre maximal d'erreurs que le codeur peut corriger par séquence.

Il a été montré par Blahut [36, 37] que la théorie des codes correcteurs d'erreurs pouvait être revisitée en utilisant une formulation et une interprétation dans le domaine spectral avec soit des corps finis soit des corps infinis tels que $\mathbb{R}$ ou $\mathbb{C}$. En particulier, un code de Reed-Solomon (RS) peut être défini de la manière suivante [36] :

Définition 2.2.3 Soit $\mathscr{F}$ un corps contenant un élément d'ordre M, le code de Reed-Solomon $(M, M-2t)$ de longueur M et dont les symboles appartiennent à $\mathscr{F}$ est l'ensemble de tous les vecteurs $\mathbf{c}$ dont le spectre (dans $\mathscr{F}$) satisfait $C_k = 0$ pour tout $k \in \{k_0+1, ...k_0+2t\}$.

Ce code est également un code linéaire par linéarité de la transformée de Fourier. Dans la suite, on considérera que $\mathscr{F} = \mathbb{C}$, où $\mathbb{C}$ est le corps des complexes. Dans ce cas là, C_k est calculé selon

$$C_k = \sum_{i=0}^{M-1} c_i W_M^{ik} \quad k = 0, 1, ..., M-1 \tag{2.72}$$

où $W_M = e^{\frac{j2\pi}{M}}$ est une racine $M^{\text{ième}}$ de l'unité dans $\mathbb{C}$. Marshall a montré [126, 127] que la plupart des propriétés des codes définis dans les corps de Galois peuvent être transposés au codage dans le corps des complexes (ou des réels). En particulier, la notion de distance de Hamming minimale garde son sens dans le corps des complexes et les algorithmes usuels de décodage des codes cycliques dans les corps finis peuvent être également employés avec des réels ou des complexes.

Le spectre d'un mot de code de Reed-Solomon défini sur un corps $\mathscr{F}$ est également à valeurs dans ce même corps $\mathscr{F}$ ($\mathscr{F} = \mathbb{C}$ ici). Par conséquent, on construit un code RS en choisissant $2t$ composantes spectrales consécutives que l'on met à 0. Le reste du vecteur, de longueur M, constitue les symboles d'information. La transformée de Fourier inverse appliquée à ce vecteur donne un mot de code RS défini dans $\mathbb{C}$.

Résultat 2.2.1 La sortie du modulateur OFDM peut être vue comme un mot de code de Reed-Solomon. En posant, $M-N = 2t$, le modulateur OFDM est un code de Reed-Solomon $(M, M-2t)$ à valeurs complexes.

Nous souhaitons ici utiliser la capacité de correction des codes RS pour compenser l'effet du canal de transmission. La capacité de correction est liée à la distance minimale du code qui n'est pas toujours facile à déterminer. Des bornes ont été établies dans la littérature afin de minorer la distance minimale du code et sa capacité de correction. Elles reposent sur des hypothèses de régularité de la répartition des zéros de fréquence dans la séquence. Dans le cas de zéros consécutifs, la borne BCH s'applique.

Notation 2.4. *On note $\mathscr{A}$ l'accord c'est à dire l'ensemble des fréquences ℓ appartenant à $\{0,1,..,M-1\}$ et telles que $C_\ell = 0$.*

Propriété 2.2.1 Si l'accord $\mathscr{A}$ contient $2t$ fréquences consécutives, alors la distance minimale du code d est telle que $d \geq 2t+1$.

Nous pouvons en déduire que ce code permet de corriger jusqu'à t erreurs. Dans l'application considérée, la perturbation est un bruit gaussien (présent sur toutes les composantes du mot de code) et un bruit impulsionnel qui n'est présent que sur quelques éléments. Nous pourrons donc utiliser le code RS pour corriger les erreurs liées au bruit impulsionnel et nous devrons adapter les algorithmes à la présence du bruit de fond gaussien.

En OFDM, l'utilisation de filtres de mise en forme analogiques réduit la taille de la bande spectrale réellement utilisable. La procédure décrite ci-dessus n'est pas directement applicable puisque seulement une petite partie des zéros consécutifs situés sur les bords du spectre est accessible. Nous proposons comme modification de remplacer les zéros hors-bande par des symboles pilotes transmis dans les systèmes OFDM pour faciliter la synchronisation et/ou l'estimation du canal. Cela a deux implications, les zéros sont remplacés par des valeurs connues et l'accord contient des fréquences qui ne sont pas nécessairement consécutives. Cette deuxième modification est la plus gênante puisqu'elle affecte directement la capacité de correction du code.

En résumé, les différences par rapport à la situation usuelle sont :

— la présence d'un bruit de fond
— l'utilisation de symboles pilotes à la place de zéros
— la position possiblement irrégulière des symboles pilotes dans la séquence

Notre contribution réside dans la prise en compte de ces spécificités. Elle sera détaillée dans la section 2.2.2. Nous rappelons maintenant les étapes de l'algorithme de Peterson-Gorenstein-Zierler spécialement conçu pour le décodage des codes BCH et des codes RS en particulier.

Algorithme de décodage des codes RS

Le premier algorithme de décodage des codes BCH a été proposé par Peterson en 1960 [163] pour les codes binaires et un an plus tard par Gorenstein et Zierler pour les codes non binaires [86]. Plusieurs simplifications ont été ajoutées ultérieurement. Nous présentons ici les étapes de l'algorithme telles qu'explicitées par Blahut dans [35]. C'est cet algorithme que nous avons choisi comme point de départ pour la correction des erreurs impulsionnelles dans un système OFDM. Nous supposons ici l'absence de bruit de fond et des syndromes consécutifs.

Notation 2.5. *On note $\mathbf{e}$ le vecteur d'erreur (ou de bruit impulsionnel dans notre application). Il est relié à la séquence reçue $\mathbf{r}$ et au mot de code $\mathbf{c}$ transmis par la relation $\mathbf{r} = \mathbf{c} + \mathbf{e}$. On utilisera une notation polynomiale qui associe à la séquence $\mathbf{e} = (i_0, i_1, ...i_{M-1})$ le polynôme $e(x) = i_0 + i_1 x + ... + i_{M-1}x^{M-1}$. On définira de manière analogue, les polynômes $c(x)$ et $r(x)$.*

Pour décrire le principe de l'algorithme de Peterson-Gorenstein-Zierler (PGZ), nous supposons que l'accord est $\mathscr{A} = \{1,2,...,2t\}$. Nous supposons également que ν erreurs se sont produites au cours de la transmission avec $1 \leq \nu \leq t$. Les syndromes sont calculés de la manière suivante :

$$S_u = r(W_M^u) = c(W_M^u) + e(W_M^u) \;\; \forall u \in \{1,2,...,2t\} \tag{2.73}$$

Par définition du code RS, on a

$$S_u = e(W_M^u) \;\; \forall u \in \{1,2,...,2t\} \tag{2.74}$$

Sachant que ν erreurs ont été commises, nous pouvons en déduire que seuls ν coefficients de $e(x)$ sont non nuls. Notons i_{f_0}, i_{f_1},...,$i_{f_{\nu-1}}$ les coefficients non nuls, on a

$$S_u = i_{f_0} W_M^{uf_0} + i_{f_1} W_M^{uf_1} + ... + i_{f_{\nu-1}} W_M^{uf_{\nu-1}} \quad \forall u \in \{1,2,...,2t\} \tag{2.75}$$

Notation 2.6. *On note* $U_l = i_{f_l - 1}$ *et* $X_l = W_M^{f_l - 1}$ *pour* $l = 1, ..., \nu$

Avec ces notations, les syndromes se réécrivent sous la forme :

$$S_u = U_1 X_1^u + U_2 X_2^u + ... + U_\nu X_\nu^u \quad \forall u \in \{1, 2, ..., 2t\} \tag{2.76}$$

Nous avons donc à résoudre un système à $2t$ équations non-linéaires et à 2ν inconnues. Les inconnues sont les couples (U_l, X_l) qui correspondent respectivement aux amplitudes et aux positions des erreurs. Nous montrons dans la suite l'unicité de la solution. Afin de résoudre le système (2.76), nous définissons au préalable des variables intermédiaires.

Définition 2.2.4 On appelle *polynôme localisateur d'erreurs*, le polynôme $\Lambda(x)$ défini selon :

$$\Lambda(x) \stackrel{def}{=} (1 - xX_1)(1 - xX_2)....(1 - xX_\nu) \tag{2.77}$$

$$\stackrel{def}{=} 1 + \Lambda_1 x + ... + \Lambda_\nu x^\nu \tag{2.78}$$

La connaissance des coefficients Λ_i pour $i \in \{1, ..\nu\}$ permet de déterminer les positions des erreurs en calculant les racines de $\Lambda(x)$ qui par définition sont égales à X_l^{-1} avec $l \in \{1, ..\nu\}$. On peut montrer que le vecteur $[\Lambda_1, \Lambda_2, ..., \Lambda_\nu]^T$ est solution du système linéaire d'équations ci-dessous :

$$\begin{bmatrix} S_1 & S_2 & \dots & S_\nu \\ S_2 & S_3 & \dots & S_{\nu+1} \\ \vdots & & & \vdots \\ S_\nu & S_{\nu+1} & \dots & S_{2\nu-1} \end{bmatrix} \begin{bmatrix} \Lambda_\nu \\ \Lambda_{\nu-1} \\ \vdots \\ \Lambda_1 \end{bmatrix} = \begin{bmatrix} -S_{\nu+1} \\ -S_{\nu+2} \\ \vdots \\ -S_{2\nu} \end{bmatrix} \tag{2.79}$$

Les conditions pour l'unicité de la solution sont données par le théorème suivant qui porte sur le déterminant de la matrice de syndromes [35].

Théorème 2.2.1 Soit $\mathbf{M}$ une matrice définie selon

$$\mathbf{M} = \begin{bmatrix} S_1 & S_2 & \dots & S_\mu \\ S_2 & S_3 & \dots & S_{\mu+1} \\ \vdots & & & \vdots \\ S_\mu & S_{\mu+1} & \dots & S_{2\mu-1} \end{bmatrix} \tag{2.80}$$

On a $\det(\mathbf{M}) \neq 0$ si $\mu = \nu$ et $\det(\mathbf{M}) = 0$ si $\mu > \nu$ où ν est le nombre effectif d'erreurs au cours de la transmission

A partir de ce résultat, la procédure de décodage peut être mise en place (algorithme 1). Elle est connue sous le nom d'algorithme de Peterson-Gorenstein-Zieler (PGZ). L'algorithme PGZ ne s'applique pas de manière directe à notre application. La section suivante détaille l'ensemble des modifications apportées pour résoudre le problème considéré ici.

2.2.2 Nos contributions

Nous nous plaçons maintenant dans la situation plus générale où : *(i)* la perturbation n'est pas seulement un bruit impulsionnel mais la somme d'un bruit impulsionnel et d'un bruit de fond, *(ii)* les syndromes ne sont pas nécessairement consécutifs, *(iii)* des symboles pilotes (ainsi que des zéros) sont utilisés pour constituer les syndromes. Dans ce cas, les composantes reçues $\{R_k\}$ s'écrivent sous la forme :

$$R_k = C_k + B_k + I_k \quad 0 \leq k \leq N - 1 \tag{2.81}$$

Algorithm 1 Peterson-Gorenstein-Zierler

$\mu \leftarrow t$ *% t est la capacité de correction*
Étape 1 : Calculer les $2t$ syndromes
Étape 2 : Calculer $\det(\mathbf{M})$
while $\det(\mathbf{M}) = 0$ **do**
  $\mu \leftarrow \mu - 1$
  Calculer $\det(\mathbf{M})$
end while *% μ est maintenant égal au nombre total d'erreurs ν*
Étape 3 : Calculer Λ_i pour $i \in \{1,..\mu\}$ à partir de (2.79)
Étape 4 : Calculer les racines de $\Lambda(x)$ et en déduire la position des erreurs
Étape 5 : Calculer U_i pour $i \in \{1,..\mu\}$ à partir de (2.76)

où $\{C_k\}$ est le symbole OFDM et où $\{B_k\}$ et $\{I_k\}$ sont respectivement les transformées de Fourier du bruit de fond gaussien et du bruit impulsionnel. Les syndromes sont donnés par

$$\begin{aligned} S_k &= R_k - C_k \quad k \in \mathscr{A} \\ &= B_k + I_k \\ &= \sum_{n=0}^{M-1} b_n W_M^{nk} + \sum_{m=0}^{\nu-1} i_{f_m} W_M^{f_m k} \end{aligned} \tag{2.82}$$

où (comme dans la section précédente) ν est le nombre total d'erreurs (impulsionnelles) lors de la transmission et $\{f_m\}_{m\in\{0,1,...\nu-1\}}$ correspond aux positions des erreurs dans la séquence. Le syndrome s'écrit donc comme une somme de sinusoïdes complexes noyées dans un bruit gaussien dont nous devons déterminer le nombre, et pour chacune d'entre elles l'amplitude et la fréquence. La particularité ici est que le nombre d'échantillons disponibles est très faible mais nous savons que les fréquences sont à valeurs entières.

① Algorithme de décodage avec syndromes non consécutifs

Nous oublions dans un premier temps la présence du bruit de fond. La procédure PGZ est équivalente à la résolution de l'équation

$$\lambda_l i_l = 0 \quad \forall l \in \{0,1,...,M-1\} \tag{2.83}$$

où $\{i_l\}$ est la séquence de bruit impulsionnel et λ_l joue le rôle d'une séquence d'annulation telle que $\lambda_l = 0$ si $i_l \neq 0$. Avec des notations vectorielles, on a

$$diag(\mathbf{i})\lambda = 0 \quad \forall l \in \{0,1,...,M-1\} \tag{2.84}$$

où $\mathbf{i} = [i_0, i_1, ..., i_{M-1}]^T$ et $\lambda = [\lambda_0, \lambda_1, ..., \lambda_{M-1}]$. Cette équation peut être exprimée à l'aide des syndromes par passage au domaine fréquentiel selon

$$\mathbf{TF}_M diag(\mathbf{i}) \mathbf{TF}_M^H \mathbf{TF}_M \lambda = 0 \tag{2.85}$$

ou de manière équivalente

$$\mathbf{S}\Lambda = 0 \tag{2.86}$$

$$\mathbf{S} = \begin{bmatrix} S_0 & S_1 & \dots & S_{M-1} \\ S_1 & S_2 & \dots & S_0 \\ \vdots & \vdots & \vdots & \vdots \\ S_{M-1} & S_0 & \dots & S_{M-1} \end{bmatrix} \tag{2.87}$$

$$\Lambda = \begin{bmatrix} \Lambda_{M-1} \\ \vdots \\ \Lambda_1 \\ \Lambda_0 \end{bmatrix} = \mathbf{T}\mathbf{F}_M\lambda \tag{2.88}$$

C'est une équation très similaire à (2.79) qui doit nous permettre de calculer Λ. Comme précédemment Λ contient les coefficients du polynôme localisateur d'erreurs, $\Lambda(x)$, défini ici selon

$$\Lambda(x) = \Lambda_0 + \Lambda_1 x + ... + \Lambda_{M-1}x^{M-1} \tag{2.89}$$

En effet, la relation (2.88) nous permet d'affirmer que $\lambda_i = \Lambda(W_M^{-i})$. Nous savons que $\lambda_i = 0$ si une erreur est survenue à la position i. Par conséquent, W_M^{-i} avec $i \in \mathscr{A}$ est une racine de $\Lambda(x)$ ce qui nous permettra de localiser les erreurs une fois les coefficients de $\Lambda(x)$ déterminés. L'algorithme de décodage est organisé en trois étapes principales comme l'algorithme PGZ :

— Calculer $\{\Lambda_i\}_{i=0}^{M-1}$ à partir de (2.86),
— Calculer les racines du polynôme localisateur d'erreurs placées sur le cercle unité,
— Connaissant les syndromes et la position des erreurs, calculer les amplitudes $\{i_\ell\}_{\ell=0}^{M-1}$.

La généralisation de la position des syndromes pose un certain nombre de problèmes. L'étape 1 doit être réadaptée : la matrice $\mathbf{S}$ contient des syndromes, S_i avec $i \in \{\mathscr{A}\}$ et des valeurs inconnues, S_i avec $i \notin \{\mathscr{A}\}$. Dans le cas consécutif, on travaille sur une sous-matrice de $\mathbf{S}$ constituée par les lignes et colonnes contenant les syndromes. Il s'agit par construction de lignes et colonnes consécutives. Les coefficients du polynôme $\Lambda(x)$ à déterminer sont également consécutifs. Par conséquent $\Lambda(x)$ est un polynôme qui a pour racine W_M^{-i} avec $i \in \mathscr{A}$ et éventuellement 0. Dans le cas non-consécutif, l'extraction de $\mathbf{S}$ d'une matrice ne contenant que des syndromes est plus difficile à trouver. Le polynôme localisateur n'a plus des coefficients non-nuls consécutifs, son degré est donc supérieur au nombre d'erreurs. Il contient W_M^{-i} avec $i \in \mathscr{A}$ pour racines mais pas seulement. Nous détaillons la résolution de (2.86) pour le cas général dans la section suivante.

② Extension des conditions sur la position des syndromes

Nous supposons dans cette partie que le nombre d'erreurs impulsionnelles est exactement égal à ν. Par conséquent, d'après (2.83), λ a au moins ν composantes égales à 0 et $M-\nu$ composantes aux valeurs arbitraires qui seront pour nous des degrés de liberté. Puisque $\Lambda = \mathbf{T}\mathbf{F}_M\lambda$, il y a également $M-\nu$ degrés de liberté sur Λ. Nous choisissons $M-\nu-1$ composantes de Λ égales à 0 et 1 composante égale à 1. Les autres composantes devront être calculées à l'aide des syndromes. Nous nous ramenons alors à une résolution de système proche de (2.79) selon :

$$\mathbf{S}_c\Lambda^{(\nu)} = 0 \tag{2.90}$$

où $\mathbf{S}_c$ est obtenue à partir de $\mathbf{S}$ en sélectionnant $\nu+1$ colonnes et où $\Lambda^{(\nu)}$ contient $\nu+1$ composantes (celles qui ne sont pas forcées à 0). Les éléments de cette sous-matrice doivent tous être connus et donc être des syndromes.

Notation 2.7. *Nous notons $\mathbf{S}^{(\nu)}$ une sous-matrice de $\mathbf{S}_c$ de taille $\nu \times \nu$ et de rang ν. Cette matrice ne contient que des syndromes.*

Dans le cas où $\mathscr{A} = \{1, 2, ..., 2\nu\}$, la matrice $\mathbf{S}^{(\nu)}$ contient les ν premières lignes et colonnes de $\mathbf{S}$ et seulement 2ν syndromes sont nécessaires pour corriger ν erreurs. Ce ne sera plus nécessairement le cas si les syndromes ne sont pas consécutifs. De plus, nous avons prouvé que, dans le cas consécutif, $\mathbf{S}^{(\nu)}$ est de rang plein. Dans le cas général, la preuve est moins évidente. Nous proposons dans la suite de cette section une condition nécessaire sur la position des syndromes pour que $\mathbf{S}^{(\nu)}$ soit de rang plein.

Propriété 2.2.2 Par construction, $\mathbf{S}^{(\nu)}$ s'écrit sous la forme

$$\mathbf{S}^{(\nu)} = \begin{bmatrix} S_{m_0+\theta_0+\delta_0} & S_{m_0+\theta_0+\delta_1} & \cdots & S_{m_0+\theta_0+\delta_{\nu-1}} \\ S_{m_0+\theta_1+\delta_0} & S_{m_0+\theta_1+\delta_1} & \cdots & S_{m_0+\theta_1+\delta_{\nu-1}} \\ \vdots & \vdots & \vdots & \vdots \\ S_{m_0+\theta_{\nu-1}+\delta_0} & S_{m_0+\theta_{\nu-1}+\delta_1} & \cdots & S_{m_0+\theta_{\nu-1}+\delta_{\nu-1}} \end{bmatrix} \tag{2.91}$$

où m_0, $\{\theta_k\}_{k=0}^{\nu-1}$, $\{\delta_k\}_{k=0}^{\nu-1}$ sont des entiers appartenant à $\{0, 1, ..., M-1\}$ tels que $(m_0 + \theta_i + \delta_j)_{0 \leq i,j \leq \nu-1} \subset \mathscr{A}$.

Nous cherchons donc une condition sur $\{\theta_k\}_{k=0}^{\nu-1}$ et $\{\delta_k\}_{k=0}^{\nu-1}$ pour que $\mathbf{S}^{(\nu)}$ soit inversible quelque soit la position des ν erreurs.

Propriété 2.2.3 Soit $\{f_i\}_{i=0}^{\nu-1}$ les positions des ν erreurs telles que $0 \leq f_0 < f_1 < ... f_{\nu-1} \leq M-1$, on a

$$\mathbf{S}^{(\nu)} = \mathbf{U}_{\theta}^{(\nu)} \mathbf{P}^{(\nu)} \left[\mathbf{U}_{\delta}^{(\nu)}\right]^H \tag{2.92}$$

$$\mathbf{U}_{\theta}^{(\nu)} = \begin{bmatrix} W_M^{f_0\theta_0} & W_M^{f_1\theta_0} & \cdots & W_M^{f_{\nu-1}\theta_0} \\ W_M^{f_0\theta_1} & W_M^{f_1\theta_1} & \cdots & W_M^{f_{\nu-1}\theta_1} \\ \vdots & \vdots & \vdots & \vdots \\ W_M^{f_0\theta_{r-1}} & W_M^{f_1\theta_{r-1}} & \cdots & W_M^{f_{\nu-1}\theta_{r-1}} \end{bmatrix} \tag{2.93}$$

$$\mathbf{P}^{(\nu)} = diag([i_{f_0} W_M^{f_0 m_0}, i_{f_1} W_M^{f_1 m_0}, \ldots, i_{f_{\nu-1}} W_M^{f_{\nu-1} m_0}]) \tag{2.94}$$

$$\mathbf{U}_{\delta}^{(\nu)} = \begin{bmatrix} W_M^{f_0\delta_0} & W_M^{f_1\delta_0} & \cdots & W_M^{f_{\nu-1}\delta_0} \\ W_M^{f_0\delta_1} & W_M^{f_1\delta_1} & \cdots & W_M^{f_{\nu-1}\delta_1} \\ \vdots & \vdots & \vdots & \vdots \\ W_M^{f_0\delta_{r-1}} & W_M^{f_1\delta_{r-1}} & \cdots & W_M^{f_{\nu-1}\delta_{r-1}} \end{bmatrix} \tag{2.95}$$

(2.96)

Sous l'hypothèse de ν erreurs subies lors de la transmission, la matrice $\mathbf{P}^{(\nu)}$ est de rang plein et la valeur de m_0 n'a pas d'influence sur le rang de $\mathbf{S}^{(\nu)}$. Comme $\{\theta_k\}_{k=0}^{\nu-1}$ et $\{\delta_k\}_{k=0}^{\nu-1}$ représentent les écarts entre les syndromes, nous pouvons en conclure que le rang de $\mathbf{S}^{(\nu)}$ est déterminé par l'espacement entre les symboles. La matrice $\mathbf{S}^{(\nu)}$ est de rang plein si $\mathbf{U}_{\theta}^{(\nu)}$ et $\mathbf{U}_{\delta}^{(\nu)}$ sont de rang plein. Nous étudions dans la suite le rang de $\mathbf{U}_{\theta}^{(\nu)}$. Les conditions trouvées sur $\{\theta_k\}_{k=0}^{\nu-1}$ s'appliqueront à l'identique sur $\{\delta_k\}_{k=0}^{\nu-1}$ pour que $\mathbf{U}_{\delta}^{(\nu)}$ soit de rang plein.

Des conditions sont déjà connues pour que la matrice soit inversible. Nous avons déjà étudié le cas consécutif, il correspond à $\theta_i - \theta_{i-1} = \delta_i - \delta_{i-1} = 1$ pour tout $i \in \{1, ...\nu - 1\}$. C'est un cas particulier de la borne BCH pour laquelle on choisit $\theta = 1$ qui inclut aussi le cas des syndromes équirépartis dans la séquence. Des conditions plus générales, ont été données par Roos, Hartmann

et Tzeng [91, 174] qui admet comme configuration $s+1$ sous-blocs contenant chacun d_0-1 syndromes uniformément espacés. Notre objectif est de trouver les conditions les plus générales possibles. Dans ce but, nous avons établi dans [5] une condition nécessaire pour que $\mathbf{U}_\theta^{(\nu)}$ soit une matrice de rang plein :

Résultat 2.2.2 Si $\mathbf{U}_\theta^{(\nu)}$ est de rang plein, les conditions suivantes sont remplies [a] :
- $\exists i_1 \in \{1,2,...,\nu-1\}$ tel que $pgcd(|\theta_{i_1}-\theta_0|,M)=1$ et $pgcd(|\theta_i-\theta_0|,M)\leq \nu-1\ \forall i\neq i_1\in\{1,...,\nu-1\}$
- $\exists i_2 \in \{2,...,\nu-1\}$ tel que $pgcd(|\theta_{i_2}-\theta_1|,M)=1$ et $pgcd(|\theta_i-\theta_1|,M)\leq \nu-2\ \forall i\neq i_2\in\{2,...,\nu-1\}$
-
- $\exists i_{\nu-2} \in \{\nu-2,\nu-1\}$ tel que $pgcd(|\theta_{i_{\nu-2}}-\theta_{\nu-3}|,M)=1$ et $pgcd(|\theta_i-\theta_{\nu-3}|,M)\leq 2$ $\forall i\neq i_{\nu-2}\in\{\nu-2,\nu-1\}$
- $pgcd(|\theta_{i_{\nu-1}}-\theta_{\nu-2}|,M)=1$

a. Les $\{\theta_i\}$ ne sont pas nécessairement classés par ordre croissant ou décroissant.

La démonstration se fait de manière récursive. Les détails de la preuve sont donnés dans [5]. En posant $\theta_k=k\theta$ et $\delta_k=k\delta$ avec $pgcd(\theta,M)=pgcd(\delta,M)=1$, on peut vérifier que les conditions données dans la borne de Hartmann-Tzeng vérifient la condition nécessaire du résultat 2.2.2.

Nous avons donné une condition nécessaire pour que la matrice $\mathbf{S}^{(\nu)}$ soit inversible. Nous devons également vérifier que le polynôme localisateur possède exactement ν racines du type $W_M^{-f_k}$ où f_k est un entier. Cette condition n'est pas garantie *a priori* puisque les degré de $\Lambda(x)$ peut être supérieur à ν dans le cas général. Une condition nécessaire et suffisante est donnée ci-dessous.

Résultat 2.2.3 Soit $\mathscr{F}=\{f_k\in\mathbb{N}:\Lambda(W_M^{-f_k})=0\}$ l'ensemble contenant la liste des positions des erreurs. On a $card(\mathscr{F})=\nu$ si et seulement si $\mathbf{U}_\delta^{(\nu+1)}$ est inversible quelque soit la position des erreurs dans la séquence.

Nous avons montré dans [1] que la condition nécessaire (2.2.2) est aussi suffisante dans le cas particuliers $\nu=2$ et $\nu=3$. Nous pourrons donc utiliser cette technique de correction de bruit impulsif à condition de vérifier que les pilotes ont des positions qui vérifient cette condition. Nous pourrons également l'appliquer à la correction du *Peak to Average Power Ratio* (PAPR) et gérer ainsi des problèmes d'écrêtement du signal.

③ Prise en compte du bruit de fond

Jusqu'à présent, nous n'avons pas considéré le bruit de fond dans notre analyse. Nous reprenons maintenant les étapes de l'algorithme PGZ en les adaptant à la présence du bruit de fond.

— **Estimation du nombre d'erreurs**. Nous supposons que la matrice $\mathbf{S}^{(r)}$ est une matrice de rang plein où $\mathbf{S}^{(r)}$ a la structure donnée dans (2.91). Nous supposons que le nombre d'erreurs impulsionnelles ν est tel que $\nu<r$ et nous écrivons $\mathbf{S}^{(r)}$ sous la forme $\mathbf{S}^{(r)}=\mathbf{I}^{(r)}+\mathbf{B}^{(r)}$ où $\mathbf{I}^{(r)}$ et $\mathbf{B}^{(r)}$ ont la même structure que $\mathbf{S}^{(r)}$ et contiennent respectivement la contribution du bruit impulsionnel et du bruit gaussien. Le bruit de fond étant supposé blanc, gaussien, de moyenne nulle et de variance σ_b^2 on a

$$E[(\mathbf{B}^{(r)})^H\mathbf{B}^{(r)}]=r\sigma_b^2 I_r \tag{2.97}$$

où I_r est la matrice identité de taille r. La différence d'amplitude entre bruit de fond et bruit impulsif fait que l'on peut estimer ν en évaluant le nombre de valeurs propres d'amplitude supérieure à $r\sigma_b^2$ [110]. En pratique, on estime ν comme étant le nombre de valeurs propres

de $(\mathbf{S}^{(r)})^H\mathbf{S}^{(r)}$ supérieures à $\phi r\sigma_b^2$ où ϕ est un coefficient multiplicatif qui permet de compenser la sous-estimation de l'estimateur due au faible nombre d'observations. On notera $\widetilde{\nu}$ l'estimation de ν.

— **Localisation et amplitude des erreurs**. La localisation des erreurs passe par la résolution d'une équation du type (2.79) et qui, dans le cas général de symboles non consécutifs, s'écrit

$$\begin{bmatrix} I_{m_0+\theta_0+\delta_0} & \cdots & I_{m_0+\theta_0+\delta_{\widetilde{\nu}-1}} \\ \vdots & \vdots & \vdots \\ I_{m_0+\theta_{r+s-1}+\delta_0} & \cdots & I_{m_0+\theta_{r+s-1}+\delta_{\widetilde{\nu}-1}} \end{bmatrix} \Lambda^{(\widetilde{\nu})} = - \begin{bmatrix} I_{m_0+\theta_0+\delta_{\widetilde{\nu}}} \\ \vdots \\ I_{m_0+\theta_{r+s-1}+\delta_{\widetilde{\nu}}} \end{bmatrix} \tag{2.98}$$

que nous noterons sous la forme compacte

$$\mathbf{I}^{(\widetilde{\nu})}\Lambda^{(\widetilde{\nu})} = \mathscr{I}^{(\widetilde{\nu})} \tag{2.99}$$

On peut obtenir une estimation $\widetilde{\Lambda}^{(\widetilde{\nu})}$ de $\Lambda^{(\widetilde{\nu})}$ par

$$\widetilde{\Lambda}^{(\widetilde{\nu})} = \left((\mathbf{S}^{(\widetilde{\nu})})^H\mathbf{S}^{(\widetilde{\nu})} - \widetilde{\nu}\sigma_b^2\mathbf{Id}\right)^{-1}(\mathbf{S}^{(\widetilde{\nu})})^H\mathscr{S}^{(\widetilde{\nu})} \tag{2.100}$$

$$\approx \left((\mathbf{I}^{(\widetilde{\nu})})^H\mathbf{I}^{(\widetilde{\nu})}\right)^{-1}(\mathbf{I}^{(\widetilde{\nu})})^H\mathscr{I}^{(\widetilde{\nu})} = \Lambda^{(\widetilde{\nu})} \tag{2.101}$$

où $\mathscr{S}^{(\widetilde{\nu})}$ est l'analogue de $\mathscr{I}^{(\widetilde{\nu})}$ dans lequel $\{I_{m_0+\theta_i+\delta_{\widetilde{\nu}}}\}$ est remplacé par $\{S_{m_0+\theta_i+\delta_{\widetilde{\nu}}}\}$. Les positions des erreurs sont alors obtenues comme les $\widetilde{\nu}$ valeurs de k donnant les plus petites valeurs de $|\widetilde{\Lambda}(W_M^{-k})|$ avec $k \in \{0,1,...,M-1\}$. Les amplitudes se déduisent par résolution d'un système analogue à (2.76) au sens des moindres carrés.

④ Ajout d'un contrôle *a posteriori*

Le bruit de fond combiné au faible nombre d'observations peut conduire à divers dysfonctionnements de l'algorithme : mauvaise estimation des positions des erreurs, mauvaise estimation du nombre d'erreurs ou dépassement de la capacité de correction. Pour augmenter la fiabilité de la méthode, nous avons proposé dans [4] d'ajouter un contrôle *a posteriori* permettant de détecter les dysfonctionnements éventuels. Ce test est basé sur la comparaison à un seuil de la norme de l'écart entre les syndromes observés et les valeurs estimées des syndromes obtenues à partir des estimations du nombre d'erreurs, de leurs positions et amplitudes. Le seuil est fixé en utilisant les lois de probabilité correspondant aux deux situations : fonctionnement correct et dysfonctionnement. Elles sont données dans [4] où le calcul du seuil est également détaillé. Un procédé similaire peut être appliqué sur la norme du vecteur de syndromes avant le décodage afin de décider si celui-ci est nécessaire.

Ultérieurement, dans [6], nous avons proposé une structure en cascade basée sur l'algorithme de décodage présenté dans cette section. Cette organisation hiérarchique réduit la complexité globale de l'algorithme et augmente sa fiabilité. Cette structure s'applique à la fois à la correction de bruit impulsif et aussi à la réduction du niveau de PAPR. Dans cette nouvelle structure, on peut obtenir une expression analytique des probabilités *a priori* et des seuils de décision augmentant la fiabilité du résultat. Ce n'était pas le cas dans [4] où les seuils étaient obtenus par simulation. Les détails de la structure en cascade ainsi que les lois de probabilités obtenues sont donnés dans [6].

⑤ Réduction du niveau de PAPR

Nous avons dit à plusieurs reprises dans ce document que le facteur de crête encore appelé PAPR pouvait être réduit de manière significative en utilisant l'algorithme de décodage présenté précédemment. Nous donnons ici quelques éléments montrant que ce problème peut être modélisé comme un problème de correction de bruit impulsionnel.

Le signal OFDM a une grande dynamique et donc un niveau PAPR important pouvant entraîner des distorsions notamment au niveau des amplificateurs. L'écrêtement (*clipping*) est un moyen

de réduire le niveau de PAPR, il est souvent produit par la saturation des amplificateurs. Cette approche conduit à une distorsion non-linéaire telle que la réapparition de pics après la conversion analogique/numérique. Notre objectif ici est de corriger les erreurs dues à l'écrêtement. Dans [95], l'effet de l'écrêtement est assimilé à un bruit impulsionnel qui peut être corrigé par des codes de Reed-Solomon dans le corps des complexes. Nous proposons ici de suivre cette approche en la généralisant à l'utilisation des symboles pilotes. L'effet de l'écrêtement est caractérisé dans [24] où il est montré que l'erreur d'écrêtement est un évènement rare que l'on peut modéliser comme un bruit impulsionnel. Cela nous permet d'utiliser le modèle de bruit décrit dans ce chapitre dans lequel la variance du bruit impulsif est donnée par l'énergie de la partie écrêtée du signal et où la probabilité d'occurrence d'un bruit impulsif est la probabilité d'écrêtement pour un niveau de PAPR fixé. Nous avons considéré les deux situation suivantes :

— PAPR calculé à la sortie du modulateur. [12]

Définition 2.2.5 — PAPR en sortie du modulateur. Le PAPR est défini selon

$$PAPR(\mathbf{c}) = \frac{\max_{k=0,1,\ldots,M-1} |c_k|^2}{E_s} \tag{2.102}$$

où E_s est la puissance moyenne du symbole et $\mathbf{c}$ représente la séquence modulée.
Le niveau d'écrêtement [a] est alors $A_c = E_s 10^{\frac{PAPR_{dB}}{10}}$.

a. A_c est le niveau de tolérance pour la dynamique du signal pour un niveau de PAPR donné, si $|c_k| > A_c$ alors le signal est écrêté.

Dans ce cas, on applique au vecteur d'observation une égalisation de type *zero-forcing*. On obtient alors une expression très similaire à (2.81) qui permet d'utiliser la méthode de correction de bruit impulsionnel [6].

— PAPR calculé à l'entrée du démodulateur. Les fortes valeurs du PAPR modélisent alors les problèmes subits par les convertisseurs ou par les amplificateurs faible bruit.

Définition 2.2.6 — PAPR en entrée du démodulateur (183). Le PAPR est défini selon

$$PAPR(\mathbf{r}) = \frac{\max_{k=0,1,\ldots,M-1} |r_k|^2}{|H_k|^2 E_s + \sigma_b^2} \tag{2.103}$$

où E_s est la puissance moyenne du symbole, $\mathbf{r}$ est la séquence reçue avant démodulation, $\{H_k\}$ la transformée de Fourier du canal et σ^2 est la variance du bruit gaussien.
Le niveau d'écrêtement est alors $A_r = (|H_k|^2 E_s + \sigma_b^2) 10^{\frac{PAPR_{dB}}{10}}$.

Dans ce cas, l'écrêtement peut être modélisé comme un bruit impulsionnel additif à la sortie du canal et la méthode de correction du bruit impulsif s'applique directement.

2.2.3 Résultats numériques

De nombreux résultats numériques ont été présentés dans [1], [5, 6]. Nous reprenons ici un exemple de correction de bruit impulsif et un exemple de réduction du PAPR dans le contexte d'HIPERLAN 2.

Conditions expérimentales 2.5 — HIPERLAN 2 (53, 113) et modèle de canal. Pour HIPERLAN 2, le nombre de porteuses est fixé à $M = 64$. Parmi elles

— 12 porteuses sont nulles (11 sur les bords et 1 au centre),
— 4 porteuses transportent des symboles connus (pilotes) que nous appellerons P_1, P_2, P_3 et P_4

12. Généralement, les non-linéarités au niveau de l'émetteur sont dues aux amplificateurs.

et qui sont situées aux positions $\{11,25,39,53\}$,
— 48 porteuses transportent les symboles d'information.

Cette norme contient 7 modes différents, nous considérerons les deux ci-dessous :
— Mode 3 : Modulation QPSK et code canal de taux $R = \frac{1}{2}$
— Mode 6 : Modulation 16QAM et code canal de taux $R = \frac{3}{4}$

Nous considérons un modèle de canal classe A. Les paramètres du bruit impulsionnel sont choisis conformément au standard ADSL pour lequel le paramètre de Bernoulli est $p = 10^{-3}$ et la variance du bruit impulsionnel est $\sigma_i^2 = (70)^2\sigma_b^2$.

Nous nous intéressons tout d'abord à la correction de bruit impulsionnel. Nous cherchons à corriger $\nu = 2$ erreurs, dans ce cas la condition nécessaire et suffisante (2.2.2) s'applique. Nous considérons deux accords différents :
— *HIPERLAN modifié* : $\mathscr{A} = \{11;32;38;53;59\}$ qui utilise P_1, P_4 et P_3 en le décalant de la position 39 à 38 ainsi que le zéro central et un zéro du bord. Avec les notations de la section 2.2.2, on a $m_0 = 0$, $\delta_0 = 11$, $\theta_0 = 0$, $\theta_1 = 21$, $\theta_2 = 27$, $\delta_1 = 32$, $\delta_2 = 38$. La capacité de correction est égale à 2 et le polynôme localisateur a deux racines du type W_M^{-k}.
— *HIPERLAN standard* : $\mathscr{A} = \{11;25;39;53;60\}$ qui utilise P_1, P_2, P_3 et P_4 et un zéro du bord. Cet accord ne vérifie pas la condition nécessaire et suffisante contrairement à l'accord précédent. En particulier, lorsque les erreurs sont espacée de 32, le polynôme localisateur a quatre racines du type W_M^{-k} au lieu des deux souhaitées. Ce problème peut être résolu en utilisant le test de contrôle *a posteriori*.

La figure 2.11 (gauche) illustre les résultats obtenus avec la méthode de correction du bruit impulsif développée dans ce chapitre et pour chacun des deux accords décrits ci-dessus. Ces résultats sont comparés à ce que l'on obtiendrait sans correction ainsi qu'à ceux obtenus en l'absence de bruit impulsionnel. On peut constater l'efficacité de la méthode qui peut être mise en oeuvre sans modification de la norme puisque les résultats du cas HIPERLAN standard sont quasiment identiques à ceux du cas HIPERLAN modifié.

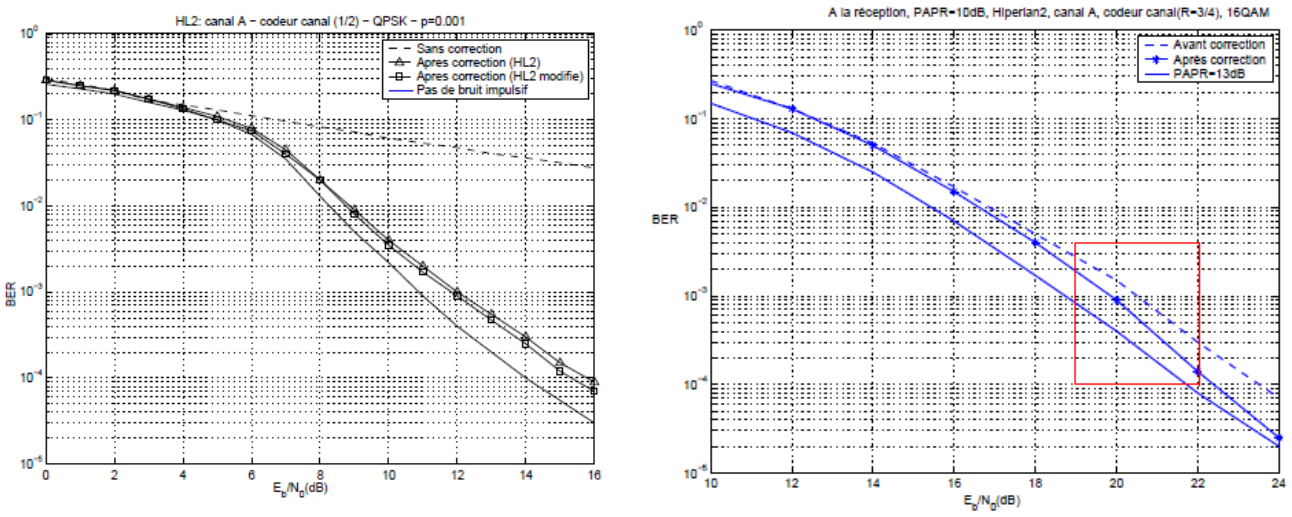


FIGURE 2.11 – Performance en terme de BER. (Gauche) : correction de bruit impulsionnel, (Droite) : réduction du niveau de PAPR.

Nous nous intéressons maintenant à la réduction du niveau de PAPR. Le PAPR est ici calculé au niveau du récepteur, il est en général de l'ordre de $13dB$. Nous souhaitons le réduire à $10dB$, l'écrêtement résultant sera alors traité comme un bruit impulsionnel à corriger. Nous considérons dans la figure 2.11 (droite), HIPERLAN 2 en mode 6. La probabilité d'occurrence d'un ecrêtement

est $p = 2 \times 10^{-3}$ et $\sigma_i^2 = 0.31$. Ces valeurs ont été obtenues par simulation. Nous comparons, en terme de BER, les performances obtenues avec un PAPR à $10dB$ suivi d'un algorithme de correction à celles obtenues avec un PAPR à $10dB$ sans correction ainsi qu'à celles obtenues lorsque le PAPR est de $13dB$. On observe un gain de $1.5dB$ pour un BER de 10^{-4} qui correspond au point de fonctionnement du mode 6.

2.2.4 Conclusion

La gestion (suppression ou diminution) d'un bruit de type impulsionnel dans des systèmes de communication fait toujours partie des sujets d'actualité comme en attestent de nombreuses publications récentes [105, 152, 167, 203]. Nous avons eu la chance d'avoir été parmi les premiers à proposer une méthode applicable à un système OFDM réaliste, aussi nos travaux sur ce sujet sont relativement bien référencés dans la littérature. La méthode proposée est évidemment perfectible et nous pouvons pointer deux éléments. Le premier est la nécessité de données structurées selon un certain modèle (voir le résultat 2.2.2) que nous avons voulu le plus général possible mais qui impose quand même des contraintes fortes sur la position des valeur connues. Par ailleurs, la méthode que nous avons proposée est sensible au bruit de fond et nous avons dû mettre en place un procédé pour gérer cette faiblesse. Parmi les méthodes récentes proposant une alternative à nos travaux, nous pouvons citer les publications [152, 167] par *Al-Naffouri et al.* qui utilisent la nature parcimonieuse du bruit impulsionnel à des fins d'estimation. Comme pour notre méthode, l'existence de symboles pilotes ou de porteuses non utilisées est nécessaire afin d'estimer le vecteur de bruit impulsionnel. L'algorithme proposé suit les mêmes étapes que notre algorithme. La différence majeure s'observe dans la méthode choisie pour estimer la position des erreurs qui est l'algorithme d'optimisation convexe proposé par *Candes et al.* [43, 44]. La solution obtenue est ensuite affinée à l'aide d'une méthode du maximum de vraisemblance *a posteriori* en introduisant un *a priori* Bernouilli-Gaussien dans le critère. Dans [152, 167], la méthode proposée est comparée à d'autres méthodes de type *compressive sampling* mais pas à notre approche ce qui ne permet pas d'évaluer la différence en terme de performances. Des méthodes d'optimisation convexe ont également été proposées dans la littérature récente pour minimiser le PAPR. Nous pouvons citer en particulier les références [10, 125].

BILAN (THÈSE/PUBLICATION)

- 1 thèse soutenue (F. Abdelkefi),
- 3 publication dans des revues internationales avec comité de lecture (*IEEE Trans. on Commun.* [R.5], [R.6] et [R.8]),
- 8 publications dans les actes de conférences internationales avec comité de lecture,
- 2 publications dans les actes de conférences nationales avec comité de lecture.

2.3 Diffusion sur les canaux sans fils

Je présente ici des résultats obtenus lors de la thèse de Z. Mheich débutée en octobre 2010 et soutenue en juin 2014 et portant sur l'étude des canaux de diffusion gaussien à deux utilisateurs. Dans cette partie, je présente une sélection de résultats qui me semble représentative du travail réalisé. Nous avons considéré un canal de diffusion gaussien dans les deux situations suivantes :

— la source envoie un message commun à destination de deux utilisateurs (ou groupes d'utilisateurs) et un message privé à destination d'un seul d'entre eux,
— la source envoie un message commun à destination des deux utilisateurs et un message confidentiel à destination du récepteur légitime.

Les limites théoriques de communication sont en général connues mais ne sont pas forcément applicables en pratique. En particulier, les régions de débits atteignables pour un canal de diffusion gaussien à deux utilisateurs ont été étudiées dans [55] puis généralisées à un nombre quelconque d'utilisateurs par Bergmans [30, 31]. Il a alors été montré que la limite théorique de la région de capacité peut être atteinte si l'alphabet d'entrée est lui même gaussien. Les alphabets utilisés en pratique sont à cardinalité fini et les différentes valeurs sont en général équiprobables. Nous souhaitons ici évaluer les régions de débits atteignables dans un contexte plus réaliste reposant sur des modulations finies. Nous considérons plusieurs stratégies de transmission : le partage de temps, la superposition de modulations et le codage par superposition. L'objectif est de déterminer le prix à payer (en terme de débit) pour l'utilisation des techniques de transmission les plus simples en terme d'implémentation. Nous souhaitons également pouvoir identifier les situations où la mise en place de techniques plus complexes vis à vis du processus de codage et/ou de décodage pourraient s'avérer indispensables.

Nous présentons ici uniquement la méthode et les résultats obtenus dans le premier contexte (message commun à deux utilisateurs et un message privé). Des aménagements ont dû être apportés pour traiter la deuxième configuration (message sécurisé vers un deuxième utilisateur) mais la façon de poser le problème reste dans le même esprit. Les résultats et la méthode utilisée dans le contexte 1 sont décrits dans **[136]** alors que ceux correspondant au contexte 2 ont été publiés dans **[137]**.

2.3.1 Le canal de diffusion gaussien

Notation 2.8. *On note :*

- X, Y_1, Y_2, *les variables aléatoires représentant respectivement l'entrée et les deux sorties du canal,*
- $\mathscr{X}$, $\mathscr{Y}_1$ *et* $\mathscr{Y}_2$ *les alphabets associés,*
- W_1 *le message privé destiné à l'utilisateur 1 uniquement,*
- W_2 *le message commun destiné aux deux utilisateurs.*

Le système considéré est constitué d'un codeur qui génère un mot de code $x^n(w_1,w_2)$ de longueur n à partir des deux messages w_1 et w_2. Les utilisateurs 1 et 2 reçoivent respectivement y_1^n et y_2^n. Le décodeur de l'utilisateur 1 fournit une estimation de (w_1,w_2) à partir de y_1^n alors que celui de l'utilisateur 2 fournit une estimation de w_2 à partir de y_2^n. Cela permet de diffuser de l'information avec deux niveaux de qualité que les utilisateurs pourront distinguer (ou pas) en fonction du canal de transmission. Nous nous limitons ici au cas d'un canal dégradé. Cela revient à supposer que l'entrée et les sorties forment un chaîne de Markov $X \to Y_1 \to Y_2$. Au lieu de considérer deux canaux parallèles, on se ramène à un canal en série dans lequel l'un des utilisateurs bénéficie de conditions de transmission plus favorables que l'autre. Le signal reçu par le récepteur ayant les moins bonnes conditions de transmission doit alors pouvoir s'écrire en fonction de celui de l'autre récepteur.

On considère donc un canal dégradé au travers duquel l'émetteur envoie les messages W_1 et W_2 avec les débits respectifs R_1 et R_2. La région de capacité du canal dégradé est donnée ci-après [57].

Théorème 2.3.1 La région de capacité du canal dégradé $X \to Y_1 \to Y_2$ est l'enveloppe convexe de tous les couples de débits (R_1,R_2) tels que

$$R_1 \leq I(X;Y1|U) \tag{2.104}$$
$$R_2 \leq I(U;Y_2) \tag{2.105}$$

pour toute distribution jointe de la forme $p(u,x)p(y_1,y_2|x)$ où U est une variable aléatoire auxiliaire de cardinalité bornée par $|U| \leq \min\{|\mathscr{X}|,|\mathscr{Y}_1|,|\mathscr{Y}_2|\}$.

Ici, l'utilisateur décode les deux messages (le message commun et le message privé). Nous nous intéresserons donc aux couples de débit (R_1+R_2, R_2). La région de capacité peut être atteinte en utilisant le codage par superposition. La variable auxiliaire U est alors le centre d'un nuage de points ou chaque point représente un mot de code. Le centre de nuage doit pouvoir être distingué par les deux utilisateurs et en particulier par l'utilisateur 2 (2.105). Sachant que l'utilisateur 2 peut distinguer le centre du nuage également puisque (2.105) est respectée et qu'il a des conditions plus favorables que l'utilisateur 2, la condition supplémentaire pour qu'il puisse distinguer les mots de code individuels est donnée par (2.104).

Dans le cas gaussien les signaux reçus s'écrivent sous la forme :

$$Y_1 = X + Z_1 \tag{2.106}$$
$$Y_2 = X + Z_2 = Y_1 + Z_2' \tag{2.107}$$

où $Z_1 \sim \mathcal{N}(0, \sigma_1^2)$, $Z_2 \sim \mathcal{N}(0, \sigma_2^2)$ et $Z_2' \sim \mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$ où σ_2^2 est la variance de Z_2 et où Z_1, Z_2' sont indépendants. Il s'agit d'un canal dégradé. Nous supposons ici une puissance limitée au niveau de l'émetteur définie selon $E[X^2] \leq P$. Les rapports signaux à bruit sont alors donnés par $SNR_i = \frac{P}{\sigma_i^2}$ avec $i = 1,2$. L'utilisateur 1 bénéficie de meilleures conditions de transmission que l'utilisateur 2 ce qui se traduit par $SNR_1 > SNR_2$. La région de capacité est définie par l'ensemble des débits R_1 et R_2 tels que

$$R_1 \leq C(\alpha SNR_1) \tag{2.108}$$
$$R_2 \leq C\left(\frac{(1-\alpha)SNR_2}{1+\alpha SNR2}\right) \tag{2.109}$$

avec $\alpha \in [0,1]$ et où $C(x) = \frac{1}{2}\log(1+x)$. La région est atteignable en utilisant le codage décrit dans [55]. On choisit 2^{nR_2} mots de code gaussiens $u^n(i)$ indépendants et identiquement distribués suivant la loi $\mathcal{N}(0,(1-\alpha)P)$. Pour chaque mot de code $u^n(i)$, on génère 2^{nR_1} mots de code gaussien satellites $v^n(j)$ de puissance αP. On ajoute les deux pour former les mots de code $x^n(i,j)$ selon $x^n(i,j) = u^n(i) + v^n(j)$. Les mots de code sont donc ici issus d'un alphabet gaussien. Nous avons voulu, dans ce travail, évaluer les débits atteignables en restreignant l'alphabet à un alphabet fini tels que ceux utilisés en pratique.

2.3.2 Schémas de transmission

Nous décrivons maintenant les schémas de transmission que nous allons comparer ainsi que les variables et degrés de liberté possibles et propres à chaque configuration. Cela se traduira dans la section suivante par la présence/absence de contraintes sur les variables d'optimisation du problème. De manière plus concrète, nous considérons ici un alphabet fini, les contraintes pourront porter sur la valeur des symboles de l'alphabet que nous pouvons choisir comme appartenant à une modulation usuelle [13] ou que nous pouvons optimiser. Le deuxième degré de liberté porte sur les probabilités des symboles $p(x)$ et aussi sur la probabilité jointe $p(u,x)$ entre la variable X et la variable auxiliaire U. On considère les stratégies de transmission dont la liste est donnée ci-après :

- S_1 Partage de temps (*Time sharing*). Les deux messages à transmettre sont transmis séparément dans des intervalles de temps disjoints. Dans chaque intervalle de temps, ce système est équivalent à un système de communication point à point. On considère une constellation M-PAM d'alphabet $\mathcal{X} = \{M-1-2(i-1), i = 1,...M\}$ et des symboles équiprobables. Cette stratégie sera nommée TS dans la suite du document.

13. Dans l'ensemble de ce travail, nous avons considéré comme modulation usuelle, la modulation M-PAM.

S_2 Superposition de modulations. Dans cette stratégie, les M points de la constellation sont obtenus par addition de deux variables aléatoires X_1 et X_2 avec $M = M_1 M_2$ où M_1 et M_2 sont à valeurs dans $\mathbb{N} \setminus \{0,1\}$ et représentent respectivement la cardinalité de X_1 et X_2. L'information commune est portée par X_2 et l'information à destination du seul utilisateur 1 est portée par X_1 ce qui conduit ici à $U \equiv X_2$. Le schéma de construction est le suivant [14] : on génère 2^{nR_2} mots de codes $u^n = x^{(2)n}(w_2)$ de longueur n suivant la loi P_U ; pour chacun de ces mots de code, on génère 2^{nR_1} mots de code satellites $v^n = x^{(1)n}(w_1)$ qui sont ajoutés pour former le mot de code $x^n = u^n + v^n$ en suivant la loi $P_{X|U}$. Lorsque P_{UX} est uniforme, le processus de codage est simple puisqu'il est séparable, les symboles obtenus sont également équiprobables comme c'est le cas dans la plupart des systèmes. Nous considérerons ici trois formes différentes de superposition de modulations [15] :

- $SM_{\mathscr{X},\overline{P_{UX}},\overline{P_X}}$ (variable à optimiser : $\mathscr{X}$),
- $SM_{\overline{\mathscr{X}},P_{UX},P_X}$ (variables à optimiser : P_{UX}, P_X),
- $SM_{\mathscr{X},P_{UX},P_X}$ (variables à optimiser : $\mathscr{X}$, P_{UX}, P_X).

On peut noter que la capacité du canal gaussien est atteignable en utilisant la superposition de modulations et deux variables indépendantes X_1 et X_2 appartenant à un alphabet gaussien. Ici X_1 et X_2, ne sont pas nécessairement indépendantes et les alphabets considérés sont de cardinalité finie.

S_3 Codage par superposition. Ce codage a été introduit par Cover dans [55] pour montrer le théorème 2.3.1 et est particulièrement adapté aux situations de diffusion [46, 79]. Le principe est le suivant : on génère 2^{nR2} mots de code de longueur n, $u^n(w_2)$, $w_2 \in \{1,2,\ldots,2^{nR2}\}$ selon $\prod_{i=1}^n p(u_i)$. Pour chaque mot de code $u^n(w_2)$, on génère 2^{nR1} mots de code indépendants $x^n(w_1,w_2)$ selon la loi de probabilité conditionnelle $\prod_{i=1}^n p(x_i|u_i(w_2))$. La variable auxiliaire U représente le centre du nuage d'information mais U n'est jamais transmis. Nous savons que $|\mathscr{U}| \leq \min\{\mathscr{X},\mathscr{Y}_1,\mathscr{Y}_2\}$, nous considérons ici le cas limite $|\mathscr{U}| = \min\{\mathscr{X},\mathscr{Y}_1,\mathscr{Y}_2\}$ pour lequel l'alphabet d'entrée est une M-PAM et les deux alphabets de sortie sont de cardinalité infinie (canaux gaussiens). La matrice P_{UX} sera donc ici une matrice de taille $M \times M$. Nous considérons encore une fois plusieurs configurations :

- $SC_{\overline{\mathscr{X}},P_{UX},\overline{P_X}}$ (variable à optimiser : P_{UX}),
- $SC_{\mathscr{X},P_{UX},\overline{P_X}}$ (variables à optimiser : $\mathscr{X}$, P_{UX}),
- $SC_{\overline{\mathscr{X}},P_{UX},P_X}$ (variables à optimiser : P_{UX}, P_X),
- $SC_{\mathscr{X},P_{UX},P_X}$ (toutes les variables doivent être optimisées).

2.3.3 Formulation du problème

La région des débits atteignables pour la stratégie TS (partage de temps) est obtenue par les deux relations ci-dessous dans lesquelles $0 \leq \alpha \leq 1$:

$$R_1 = \alpha \overline{R_1} \tag{2.110}$$

$$R_2 = \alpha \overline{R_2} \tag{2.111}$$

où $\overline{R_1}$ et $\overline{R_2}$ sont les débits atteignables pour un canal point à point avec une constellation M-PAM et des rapports signaux à bruit respectifs SNR_1 et SNR_2.

Pour les deux autres stratégies, le problème d'optimisation est formulé de la manière suivante :

$$\max_{P_{UX} \in \mathscr{C},\mathscr{X}} \theta \cdot I(X;Y_1|U) + (1-\theta) \cdot I(U;Y_2) \tag{2.112}$$

14. Par construction, la matrice P_{UX} dans le cas de la superposition de modulations est de taille $M_2 \times M$ et contient M éléments à calculer, les autres étant égaux à 0.

15. Dans la nomenclature choisie, les variables sur-lignées correspondent à des variables fixées aux valeurs standards (M-PAM, probabilité uniforme), les variables non-surlignées sont à optimiser.

Initialisation 0	$s \leftarrow s^{(0)}$	
Étape k	Initialisation 0	$\mathscr{X} \leftarrow \mathscr{X}^{(0)}$ où $\mathscr{X} = (x_0, x_1, .., x_{M-1})$
	Étape ℓ	$P_{UX}^{(\ell)} = \arg\max_{P_{UX} \in \mathscr{C}} L(P_{UX}, \mathscr{X}^{(\ell-1)}, s^{(k-1)})$ $(P1)$ $\mathscr{X}^{(\ell)} = \arg\max_{\mathscr{X}} L(P_{UX}^{(\ell)}, \mathscr{X}, s^{(k-1)})$ $(P2)$
	Test d'arrêt	$\lvert L(P_{UX}^{(\ell)}, \mathscr{X}^{(\ell)}, s^{(k)}) - L(P_{UX}^{(\ell-1)}, \mathscr{X}^{(\ell-1)}, s^{(k-1)}) \rvert \leq \varepsilon_L$
	$s^{(k)} = [s^{(k-1)} - \beta(P - \sum_{i,j} p_{ij}^*(s^{(k-1)}) \cdot (x_j^*(s^{(k-1)}))^2)]^+$ où $[.]^+ = \max(., 0)$	
Test d'arrêt	$\lvert s^{(k)} - s^{(k-1)} \rvert \leq \varepsilon_s$	

TABLE 2.3 – Solution numérique de résolution de (2.112)

L'ensemble $\mathscr{C}$ est un ensemble de contraintes qui contient dans la version minimale la contrainte de positivité sur chaque élément de P_{UX} ainsi que la contrainte faite à la somme des éléments d'être égale à 1, la contrainte de puissance en entrée du canal est également toujours présente. En fonction de la stratégie considérée peut s'ajouter une contrainte d'équiprobabilité sur P_X. Lorsque la stratégie choisie suppose que la constellation est une M-PAM, $\mathscr{X}$ n'est plus une variable d'optimisation. De la même manière, si P_{UX} est supposée uniforme, P_{UX} n'est plus une variable à optimiser. La somme pondérée dans (2.112) permet de maximiser le débit associé à l'information commune en choisissant $\Theta = 0$, à l'autre extrême choisir $\Theta = 0.5$ conduit à la maximisation du débit propre à l'utilisateur 1. Nous considérerons ici l'ensemble des situations intermédiaires avec $0 \leq \Theta \leq 0.5$.

Notation 2.9. *On note p_{ij} les éléments de la matrice P_{UX} tels que $p_{ij} = \Pr\{U = u_i, X = x_i\}$ avec $j \in \{0, .., M-1\}$ et $i \in \{0, .., |\mathscr{U}| - 1\}$.*

Les expressions de $I(X;Y_1|U)$ et $I(U;Y_2)$ se mettent alors sous la forme :

$$I(X;Y_1|U) = \sum_{i,j} \int_{-\infty}^{+\infty} p_{ij} P_{Y_1|X}(y_1|x_j) \log_2 \frac{(\sum_{j'} p_{ij'}) P_{Y_1|X}(y_1|x_j)}{\sum_{j'} p_{ij'} P_{Y_1|X}(y_1|x_{j'})} dy_1 \tag{2.113}$$

$$I(U;Y_2) = \sum_{i} \int_{-\infty}^{+\infty} (\sum_{j} p_{ij} P_{Y_2|X}(y_2|x_j)) \log_2 \frac{\sum_{j'} p_{ij'} P_{Y_2|X}(y_2|x_{j'})}{(\sum_{j'} p_{ij'})(\sum_{i',j'} p_{i'j'} P_{Y_2|X}(y_2|x_{j'}))} dy_2 \tag{2.114}$$

où les lois de probabilité conditionnelles $P_{Y_i|X}(y_i|x_j))$ pour $i \in \{1, 2\}$ se déduisent du caractère gaussien du canal. Pour résoudre le problème (2.112) nous le mettons sous la forme :

$$L(P_{UX}, x_0, .., x_{M-1}, s) = \theta \cdot I(X;Y_1|U) + (1-\theta) \cdot I(U;Y_2) + s \cdot (P - \sum_{i=0}^{|\mathscr{U}|-1} \sum_{j=0}^{M-1} p_{ij} \cdot x_j^2) \tag{2.115}$$

Une méthode itérative de résolution est utilisée en suivant les étapes données dans la table 2.3. Le problème (P_1) est résolu à l'aide de l'algorithme de Blahut-Arimoto en s'inspirant de la méthode utilisée dans [201]. Nous apportons des modifications à cet algorithme pour prendre en compte la contrainte de puissance et la contrainte d'équiprobabilité des symboles émis lorsqu'elle doit s'appliquer. Les conditions de convergence de l'algorithme de Blahut-Arimoto dans le cas du problème [16] (P_1) sont donnée dans le théorème 2 de [201]. Considérons maintenant le problème (P_2). La fonction $L(P_{UX}^{(\ell)}, \mathscr{X}, s^{(k-1)})$ n'est pas concave sur $\mathbb{R}^M$. Elle est en revanche concave sur un domaine $\mathscr{D}$ défini selon $\mathscr{D} = \{\mathscr{X} \in \mathbb{R}^M : |x_i - x_j| > d \ \ \forall i, j \in \{0, .., M-1\} \text{ and } i \neq j\}$ et où d

16. $I(X;Y_1|U)$ est une fonction concave en P_{UX} alors que $I(U;Y_2)$ s'écrit comme une différence de fonctions concaves en P_{UX}.

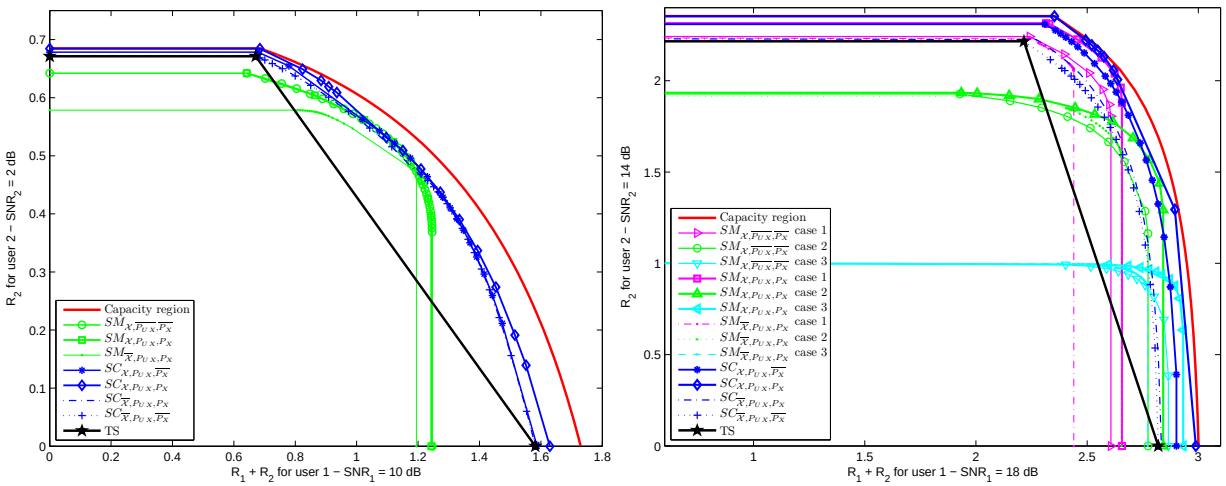


FIGURE 2.12 – Régions de débits atteignables : (gauche) $M = 4$ et $(SNR_1, SNR_2) = (10dB, 2dB)$, (droite) $M = 16$ et $(SNR_1, SNR_2) = (18dB, 14dB)$.

est un paramètre qui dépend de la taille de la constellation et du *SNR*. Ce domaine correspond à des solutions non-dégénérées. Nous utilisons ici une méthode du simplexe avec une initialisation choisie dans $\mathscr{D}$. Chacune de ces deux étapes permet d'augmenter la fonction d'objectif à chaque itération. On note $(p^*_{i,j}(s), x^*_j(s))$, $0 \leq j \leq M-1$, $0 \leq i \leq |\mathscr{U}|-1$ le point stationnaire ainsi obtenu pour une valeur s fixée. La valeur de s est gérée par la procédure itérative :

$$s^{(k+1)} = \left[s^{(k)} - \gamma \cdot \left(P - \sum_{i=0}^{|\mathscr{U}|-1} \sum_{j=0}^{M-1} p^*_{ij}(s^{(k)}) \cdot (x^*_j(s^{(k)}))^2 \right) \right]^+ \tag{2.116}$$

où $[.]^+ = \max(.,0)$. La valeur de s est ainsi augmentée ou diminuée en fonction de la valeur relative de la puissance calculée à partir de $(p^*_{i,j}(s), x^*_j(s))$ et relativement à la consigne P. On peut constater que $L(P_{UX}, \mathscr{X}, s)$ est le Lagrangien du problème (2.112). L'ensemble des étapes (P1) et (P2) et les itérations successives constituent une méthode itérative de résolution du problème

$$f(s) = \max_{P_{UX}, x_0, .., x_{M-1}} L(P_{UX}, x_0, .., x_{M-1}, s) \tag{2.117}$$

La fonction $f(s)$ est convexe, le problème d'optimisation dual $\min_{s.t.\ s\geq 0} f(s)$ dans (2.116) est résolu par une méthode de type gradient.

2.3.4 Résultats

A l'aide de l'algorithme précédent, les régions de débits atteignables ont pu être tracées pour différentes stratégies de transmission, différentes tailles de constellation et divers couples de SNR. Nous montrons dans la figure 2.12 deux exemples de résultats obtenus [17]. L'ensemble des courbes ainsi que les indicateurs utilisés pour quantifier les différences entre les stratégies sont disponibles dans [136]. Nous donnons ici simplement les conclusions principales auxquelles nous sommes parvenus. Nous les organisons autour des trois questions suivantes.

① Parmi les trois stratégies de superposition de modulations testées, peut-on en éliminer certaines ?

Nous avons observé que $SM_{\mathscr{X},\overline{P_{UX}},\overline{P_X}}$ et $SM_{\overline{\mathscr{X}},P_{UX},P_X}$ présentent en général des résultats très proches. La stratégie $SM_{\mathscr{X},\overline{P_{UX}},\overline{P_X}}$ est beaucoup plus simple à mettre en oeuvre que $SM_{\overline{\mathscr{X}},P_{UX},P_X}$ par

17. *Case 1* correspond à la stratégie de superposition de modulations telle que $M_1 = 2$ et $M_2 = 8$, *Case 2* correspond à $M_1 = 4$ et $M_2 = 4$ et *Case 3* à $M_1 = 8$ et $M_2 = 2$

conséquent $SM_{\overline{\mathscr{X}},P_{UX},P_X}$ semble peu utile. En revanche, $SM_{\mathscr{X},P_{UX},P_X}$ donne des gains supérieurs à ceux de $SM_{\mathscr{X},\overline{P_{UX}},\overline{P_X}}$ en particulier lorsque l'écart entre les SNR des utilisateurs est faible. La stratégie $SM_{\mathscr{X},P_{UX},P_X}$ qui correspond donc à la situation où l'ensemble des variables ont été optimisées n'est pas intéressante lorsque les deux utilisateurs ont des SNR très différents. Lorsque l'alphabet d'entrée est autorisé à être de cardinalité infinie, la région de capacité s'obtient en additionnant deux variables gaussiennes indépendantes. L'analyse de la matrice P_{UX} montre que lorsque l'alphabet est de taille finie, les deux variables X_1 et X_2 ne sont pas indépendantes en général.

② Que choisir entre partage de temps et superposition de modulations ?

La superposition de modulation est une meilleur option que le partage de temps lorsque l'écart de SNR est important entre les utilisateurs. Le partage de temps est quand à lui optimal dans la région où l'on cherche à maximiser le débit R_2 de l'utilisateur 2 et quand celui-ci bénéficie d'un bon rapport signal à bruit. En effet, le débit atteint pour un seul utilisateur dans la stratégie de partage de temps est obtenu en utilisant la constellation M-PAM intégralement alors qu'elle est partagée entre les utilisateurs dans la stratégie de superposition de modulation.

③ Le codage par superposition a-t-il un intérêt ?

La part de la région maximale atteignable avec le codage par superposition mais non-atteignable avec le partage de temps ou la superposition de modulations est d'autant plus important que la taille M de la modulation est faible. En effet, lorsque M augmente, le nombre de configurations différentes possibles avec la superposition de modulations augmente aussi et nous savons que lorsque M tend vers l'infini, la superposition de modulation est le schéma optimal. Le codage par superposition semble devoir être réservé aux situations avec modulation de faible cardinalité et pour des écarts de SNR importants entre utilisateurs. on constate également que dans la stratégie de codage par superposition l'optimisation des positions des symboles a peu d'influence.

④ Etude d'un *scénario*

Avant de terminer cette section, nous proposons une étude de cas. Nous étudions un système initialement constitué d'un ou de plusieurs utilisateurs (nous le(s) nommerons utilisateur 2 en reprenant les mêmes notations que précédemment). L'utilisateur 2 est à une distance d_2 de l'émetteur et atteint un débit R_0 lorsque les données sont modulées avec une 4-PAM pour laquelle les symboles sont équiprobables. Un (plusieurs) nouveaux utilisateur(s) doivent également être pris en considération. Nous les nommerons utilisateur 1, ils bénéficient d'un rapport signal à bruit meilleur que celui de l'utilisateur 2. On suppose ici que le SNR est proportionnel à l'inverse de la distance utilisateur/émetteur élevée au carré.

Conditions expérimentales 2.6 Le rapport signal à bruit de l'utilisateur 2 est $SNR_2 = 10dB$. Le débit vers l'utilisateur 2 est $R_0 = 1.582 bits/ch.use$. Initialement, seul l'utilisateur 2 est présent, le débit est R_0 et la modulation est une 4-PAM avec symboles équiprobables. L'objectif est de servir l'utilisateur 1 tout en maintenant un débit R_0 vers l'utilisateur 2 en utilisant une modulation 16-PAM optimisée. On mesure le rapport d_1/d_2 où d_i est le diamètre de la zone de couverture de l'utilisateur i.

Les résultats sont montrés sur la figure 2.13 où est tracée l'évolution de d_1/d_2 en fonction du débit atteignable par l'utilisateur 1. On peut voir par exemple que si l'utilisateur 1 est à mi-chemin entre l'émetteur et l'utilisateur 2 la stratégie de codage par superposition n'est pas nécessaire. Une stratégie de superposition de modulations est suffisante. Le codage par superposition s'avèrera beaucoup plus utile pour des situations où l'utilisateur 1 est plus proche de l'émetteur ($d_1/d_2 = 0.2$ par exemple). Dans tous les cas, la superposition de modulations conduit à des gains importants comparativement au partage de temps.

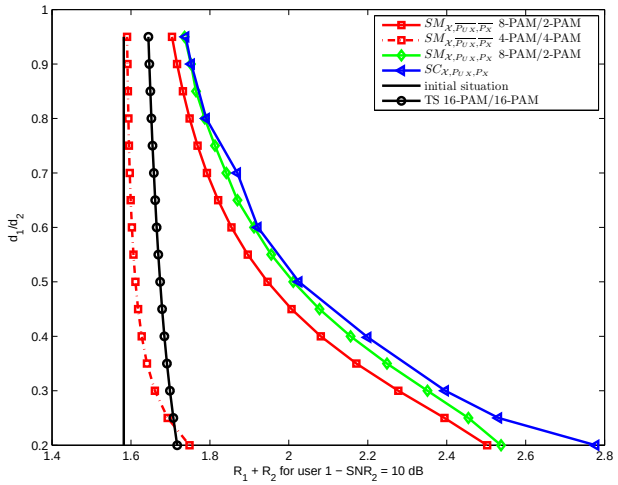


FIGURE 2.13 – Ratio d_1/d_2 en fonction de $R_1 + R_2$

2.3.5 Conclusion

La contribution principale de ce travail est d'avoir mené une étude sur les débits atteignables pour un canal gaussien à deux utilisateurs en imposant au système un principe de "réalité" reposant sur l'utilisation de modulations finies. Nous avons pu constater que l'optimalité n'était pas obtenue pour les mêmes schémas de transmission en modulation finie et infinie. Nous avons pu également donner des recommandations sur la stratégie la plus efficace ou réalisant le meilleur compromis efficacité/complexité en fonction des paramètres du système qui sont essentiellement la taille de la modulation et les valeurs des rapports signaux à bruit des utilisateurs. Nous avons également montré lors de la thèse de Zeina Mheich qu'un formalisme très similaire permettait d'obtenir les régions de débit atteignables avec une contrainte de sécurité. Ce travail pourrait être étendu assez facilement à d'autres types de modulations et généralisé vers d'autres types de canaux de transmission. Par le passage de modulations infinies vers des modulations finies, nous avons fait un pas vers une implémentation pratique toutefois la génération de symboles non équiprobables n'est pas toujours simple à réaliser en pratique. On trouve dans la littérature récente des travaux portant sur cette question pour des canaux points à point utilisant des turbo-codes [188] ou des codes LDPC [107, 189]. Nous avons également étudié lors de la thèse de Zeina Mheich un schéma adaptatif de communication sécurisée basée sur les protocoles HARQ dans lequel le débit est adapté en fonction de l'information transmise par le récepteur légitime via un canal de retour [138]. Ce travail n'a pas été présenté ici, nous en dirons quelques mots dans le chapitre suivant.

BILAN (THÈSE/PUBLICATION)

- 1 thèse soutenue (Z. Mheich),
- 2 publication dans des revues internationales avec comité de lecture (*IEEE Trans. on Commun.* [R.2], *EURASIP Journal on Wireless Communications and Networking* [R.4]), 1 article soumis à *IEEE Trans. on Commun.*.
- 3 publications dans les actes de conférences internationales avec comité de lecture.

3 — Projet de recherche

Dans ce chapitre, je présente mes perspectives de recherche à plus ou moins longue échéance. Les thèmes principaux s'articulent autour des techniques de codage (et de décodage) d'une part et également autour des techniques d'estimation adaptative/itérative et sont donc une suite logique aux travaux présentés dans le chapitre précédent. Le premier thème exposé dans la section 3.1 décrit un sujet de recherche qui a débuté en 2014 avec la thèse de Hong Nhat Nguyen et porte sur des techniques pragmatiques de coopération dans les réseaux sans fil. Le sujet développé dans la section 3.2 fait suite aux travaux menés dans la dernière partie de la thèse de Z. Mheich. Les travaux envisagés utilisent également des résultats développés dans la deuxième partie de ce manuscrit. Le troisième thème utilise des compétences et des connaissances exploitées dans les recherches présentées dans la section 2.1. Comme pour les travaux déjà réalisés, ce projet maintient un équilibre entre des recherches appliquées au domaine des télécommunications et le développement de techniques ou d'algorithmes de traitement du signal.

3.1 Stratégies de coopération dans un environnement sans fil

3.1.1 Contexte

Ce premier axe de recherche a été initié avec la thèse de H. N. Nguyen débutée en avril 2014 et co-encadrée avec P. Duhamel. Ce thème fait également l'objet d'un projet de recherche collaboratif avec l'université de Louvain dans le cadre de Newcom♯.

Nous nous intéressons à des stratégies de coopération de type relayage dans lesquelles l'intégrité des signaux reçus est protégée à l'aide d'un code détecteur d'erreur. Diverses situations de relayage ont été étudiées dans la littérature et s'articulent autour de 4 catégories principales distinguées par l'action demandée au relais et qui sont les techniques *Amplify and Forward* [116], *Decode and Forward* [135], *Demodulate and Forward* [56] et *Compute and Forward* [156]. Le relais retransmet ensuite tout ou partie de l'information reçue. En terme de support de transmission, on considère souvent, dans la littérature, des canaux orthogonaux ce qui correspond à la situation ou les communications sont séparées soit temporellement soit fréquentiellement. C'est évidemment la situation idéale vis à vis de l'interférence mais c'est aussi une situation qui demande des ressources additionnelles. De manière duale, on peut considérer des canaux à interférence dans lesquels tout ou partie de l'information envoyée par le relais se superpose à une transmission existante [176]. Le relais peut apporter un soutien à plusieurs utilisateurs en combinant linéairement (*network coding*)

les messages en provenance de plusieurs sources [11].

3.1.2 Travail envisagé

Une situation à plusieurs utilisateurs est considérée ici ; chacun transmettant des données protégées par un code canal. Le relais utilise la stratégie *demodulate and Forward* et procède à une combinaison des messages des deux utilisateurs. Les stratégies mixtes entre canaux orthogonaux et canaux à interférence seront étudiées. D'un point de vue applicatif, il s'agit de considérer la liaison montante pour un système de communication sans fil dans lequel chaque utilisateur dispose d'une ressource propre pour transmettre les données (canaux orthogonaux) alors que le relais transmet sur une ressource déjà existante sans se voir allouer un canal spécifique (canal à interférence). Pour les deux utilisateurs, le relayage est "transparent" dans le sens où il ne leur est pas demandé d'adaptation particulière de leur mode de transmission. En revanche, le récepteur doit être modifié au niveau de la station de base afin de prendre en compte cette nouvelle configuration.

Le travail à mettre en oeuvre se situe à plusieurs niveaux. On s'intéressera dans un premier temps au choix de la configuration de transmission ainsi qu'à l'optimisation du code. En effet, un certain nombre de travaux ont montré l'importance du codage canal dans une situation de relayage en particulier vis à vis de la diversité. C'est le cas par exemple de [190] qui montre que l'utilisation d'un code convolutif performant permet d'obtenir une diversité maximale en procédant à un relayage partiel préservant ainsi une partie importante des ressources. D'autres travaux [11] ont été menés pour des codes LDPC afin de permettre d'atteindre une diversité maximale tout en transmettant avec un taux de codage [1] maximal. Dans cette contribution, le relais décode et transmet et le taux de codage maximal est obtenu en poinçonnant une partie des bits de parité. Cette configuration est transposable à une situation où le relais ne décode pas mais démodule seulement au prix d'une interférence accrue. On dispose donc de divers degrés de liberté (type de code, configuration, poinçonnage,...) à évaluer et à optimiser conjointement.

Le choix du relais est aussi une question à analyser. C'est un problème qui a été beaucoup étudié pour choisir un ou plusieurs relais associés à un couple source-destination [101]. Nous sommes ici dans une situation un peu différente puisque un relais est en soutien de plusieurs utilisateurs, il y a donc un compromis à choisir.

Enfin, un travail sera à mener sur le récepteur et sur le décodage itératif plus particulièrement afin de prendre en compte les spécificités de la transmission liées aux différents rapports signal-bruit (canal source-destination et relais-destination), à l'utilisation du *network coding* et aux différents niveaux de protection selon les bits. Plusieurs structures de décodage sont possibles et devront être analysées.

3.2 Sécurité couche physique

3.2.1 Contexte

La sécurité au niveau de la couche physique [39] est un ensemble de méthodes qui consiste à utiliser les imperfections du canal de transmission (interférences, bruit, canaux à évanouissement,...) afin de garantir la confidentialité d'une communication vers un utilisateur légitime. Plusieurs résultats issus de la théorie de l'information montrent qu'il est possible de sécuriser des transmissions en présence d'un espion passif en s'aidant des imperfections du canal et aussi en utilisant des codes correcteur d'erreurs. Le partage d'une clé secrète entre l'émetteur et l'utilisateur n'est pas forcément nécessaire. Toutefois des méthodes mixtes peuvent être envisagées afin de renforcer la sécurité. Nous considérons le schéma classique de communication connu sous le nom de *wiretap channel*

1. On entend ici le taux de codage équivalent pour l'ensemble de la transmission c'est à dire le nombre total de bits de message à transmettre par les deux utilisateurs rapporté au nombre totale de bits transmis par les utilisateurs et par le relais.

[197] et que nous traduirons ici par canal à écoute. Ce modèle fait intervenir deux utilisateurs légitimes, Alice et Bob, qui échangent des informations sur le canal principal ainsi qu'un troisième individu, l'espion, qui reçoit également le message d'Alice sur un canal séparé. L'enjeu est de rendre inintelligible le message d'Alice pour l'espion qui ne bénéficie d'aucune restriction quand à sa capacité de calcul. On suppose également que l'espion connaît, au moins partiellement, le type de codage utilisé et est capable de mettre en oeuvre un décodage approprié. La sécurité doit donc être garantie par le biais des caractéristiques des canaux qui sont différents pour les deux utilisateurs, par le type de codage ou par une mise en forme particulière des signaux envoyés qui serait plus favorable à l'utilisateur légitime [89].

3.2.2 Travaux déjà entrepris

Un travail sur le thème de la sécurité au niveau de la couche physique a déjà été mené lors de la thèse de Z. Mheich. Nous l'avons abordé de deux manières différentes. Dans un premier temps, nous avons calculé les régions de débits atteignables pour un canal de diffusion avec un message commun pour deux utilisateurs et un message secret pour l'un d'entre eux. Comme pour le travail présenté dans la section 2.3, notre contribution a été de montrer l'impact, sur la région des débits atteignables, des diverses stratégies de transmission combinées à une hypothèse d'alphabet fini. Les résultats obtenus ont été publiés dans **[137]**. Dans une deuxième contribution, nous nous sommes intéressés au canal à écoute en supposant un canal à évanouissement par bloc et en supposant que l'émetteur n'a pas connaissance de l'état instantanée du canal. Nous nous sommes intéressés à un schéma de communication à redondance incrémentale basé sur un protocole HARQ (*Hybrid Automatic Retransmission Request*). Nous avons montré comment le débit de la source peut être adapté (ici en fonction de l'état passé du canal entre émetteur et utilisateur légitime) afin de garantir la sécurité et aussi l'intelligibilité du message pour le décodeur légitime [138]. Dans les travaux susmentionnés, le problème a été abordé sous l'angle de la théorie de l'information dans le sens où nous avons considéré des symboles appartenant à un alphabet gaussien, un codage aléatoire permettant d'atteindre la capacité [197] et un décodage par séquences typiques. Je propose ici d'aborder ce problème sous un angle pratique.

3.2.3 Travail envisagé

Le premier point sur lequel se positionner est le choix de la métrique permettant d'évaluer la (non)-satisfaction de la contrainte de sécurité. On utilise traditionnellement des métriques à base d'information mutuelle telles que (3.1-3.2) qui sont respectivement la contrainte de sécurité au sens faible et au sens fort

$$\lim_{n\to\infty} \frac{1}{n} I(X;Z^n) = 0 \tag{3.1}$$

$$\lim_{n\to\infty} I(X;Z^n) = 0 \tag{3.2}$$

où X est le message et où Z^n représente l'observation pour l'espion, n est la longueur du mot de code. Le choix de la métrique est lié au niveau de sécurité souhaité qui dépend lui-même de l'application. Ces deux mesures du niveau de sécurité sont asymptotiques vis à vis de la longueur du bloc et ne sont pas nécessairement adaptées à l'analyse de système pour lesquels les blocs sont de longueur finie. Des alternatives ont vu le jour, dans lesquelles la notion de sécurité est assouplie, on peut citer par exemple [114] qui applique des seuils aux taux d'erreur moyens pour évaluer la sécurité et également la fiabilité du message reçu par le récepteur légitime.

Le deuxième axe s'articule autour de la construction de codes et/ou sur le choix des paramètres d'un code permettant d'assurer les deux critères antagonistes que sont la sécurité et la fiabilité. Parmi l'ensemble des codes possibles, les codes LDPC et les *polar* codes [21] ont été étudiés par certains auteurs pour la sécurité, les turbo-codes également dans une moindre mesure. De nombreux

travaux sécurisent la communication en utilisant le poinçonnage [74, 114] pour mettre en difficulté l'espion. Cela a pour effet de réduire également la capacité de décodage du récepteur légitime. De manière totalement symétrique, on peut au contraire utiliser des codes à insertion [199]. Le principe des codes à insertion est l'introduction de bits connus, non nécessairement transmis, dans la séquence d'information en entrée du codeur. Dans un contexte de confidentialité, la valeur et la position des bits insérés sont connus du seul utilisateur légitime lui donnant un avantage certain sur l'espion. Une première étude d'une combinaison de codage par insertion et de poinçonnage a été menée récemment dans [109] en s'appuyant sur les EXIT Charts. Pour être complètement adaptative la méthode nécessite la connaissance à l'émetteur du rapport signal à bruit pour chacun des deux canaux. Ce que nous proposons ici est d'étudier la combinaison d'un codage par insertion avec du poinçonnage en remplaçant les EXIT-Charts traditionnels (*ie* utilisant l'information mutuelle $I(L;X)$) par un EXIT-Chart à courbe unique utilisant l'information mutuelle entre extrinsèques (voir section 4.6.3). Nous pourrons ainsi calculer hors-ligne les paramètres optimaux pour garantir sécurité et fiabilité à l'utilisateur légitime. En utilisant un protocole de type HARQ, l'information mutuelle entre extrinsèques au niveau du récepteur légitime pourra être transmise à l'émetteur par un canal à retour qui saura alors comment choisir les bons paramètres pour une retransmission ou un complément de transmission. C'est une façon de mettre en oeuvre de manière concrète le procédé de redondance incrémentale **[138]**. La compatibilité de la méthode proposée avec les systèmes sans fil déjà déployés (LTE, relayage, services ou communications mobile-à-mobile,...) devra ensuite être étudiée.

Les techniques de codage pour la sécurité sont plus efficaces lorsqu'elles sont combinées avec des techniques de traitement du signal permettant d'accroître les différences entre les caractéristiques du canal principal, que l'on espère le meilleur possible, et celles du canal de l'espion. Les techniques envisageables sont les techniques de filtrage spatial (*beamforming* ou *precoding*) éventuellement combinées avec l'adjonction d'un bruit artificiel [84, 112]. Ce dernier axe consistera donc à étudier de manière conjointe les techniques de codage/décodage et les techniques de traitement et de mise en forme des signaux transmis.

Dans les pistes que nous venons de citer, l'espion est supposé agissant seul et bénéficiant d'un nombre limité d'observations et d'un mauvais canal. Mais, si l'on imagine plusieurs espions avec des observations indépendantes et/ou en capacité de collaborer, les hypothèses usuelles (et que nous avons faites également ici) sont susceptibles de s'effondrer. Je m'intéresserai également à cette situation en commençant avec une structure incluant un canal à retour comme proposé plus haut. L'intérêt d'un canal à retour dans ce contexte a en effet été montrée dans [93] ; la mise en oeuvre pratique reste toutefois à explorer.

3.3 Optimisation pour les problèmes à grande dimension

3.3.1 Contexte

Le troisième axe de recherche est du domaine de l'optimisation pour les problèmes de grande dimension. C'est une thématique en plein essor en traitement du signal et qui peut être appliquée à divers domaines tels que l'imagerie médicale, la bio-informatique, ou encore la géophysique. C'est également une problématique qui touche le domaine des télécommunications que ce soit pour les réseaux de téléphonie mobile ou pour les réseaux de distribution électrique pour lesquels un volume important de données est généré et pourrait être traité et analysé en temps réel afin d'optimiser les ressources disponibles.

Les propriétés que doivent partager les solutions algorithmiques pour les problèmes à grande dimension sont une complexité calculatoire très faible, la possibilité d'un calcul parallèle ou distribué, la capacité à traiter les données au fil de l'eau. De nombreux algorithmes ont été développés ou remis au goût du jour incluant les algorithmes adaptatifs, distribués, les méthodes stochastiques pour

résoudre des problèmes convexes et non-convexes [45, 179]. Les méthodes proposées sont souvent très basiques (méthodes du premier ordre, gradient stochastique) et reposent sur des principes simples mais relativement efficaces [45].

3.3.2 Travail envisagé

Parmi l'ensemble des techniques d'optimisation possibles, je souhaite m'intéresser aux techniques d'inférence bayesienne qui peuvent être déclinées en des méthodes stochastiques (MCMC) ou en des méthodes déterministes telles que les méthodes du type bayésien variationnel. Nous avons déjà discuté dans la section 2.1 des avantages et inconvénients des méthodes stochastiques vs déterministes. Les méthodes stochastiques ont en général l'avantage de conduire à une estimation de bonne qualité mais au prix d'une complexité calculatoire plus grande. Dans une première approche, il me semble que l'on doit choisir des solutions à faible complexité ce qui plaiderait plutôt en faveur de méthodes déterministes. La plupart des implémentations existantes ne sont pas adaptées à des problèmes à grande dimension car chaque remise à jour est dépendante de l'ensemble des données. Il est donc nécessaire de proposer des solution telles que la remise à jour des estimateurs puisse être faisable à partir d'un sous-ensemble des données et de manière distribuée. Une implémentation de type *expectation propagation* [2] [141, 194] est un début de réponse à ce problème dans la mesure où l'estimation est réalisée de manière locale avec des informations transmises d'un point à un autre par propagation de messages ou de croyances. D'autres alternatives ont vu le jour dans la littérature récente en terme d'implémentation distribuée [33, 177] et offrent un premier ensemble de réponses. Les directions à explorer sont donc la mise au point de techniques d'estimation distribuées et utilisant des implémentations itératives. Cela conduit à un certain nombre de questions. L'ordonnancement par exemple est un point important pour lequel la masse considérable de données peut conduire à des choix différents des choix traditionnels. La convergence des méthodes itératives n'est pas toujours garantie ; c'est évidemment un point qui nécessitera une attention particulière. La vitesse de convergence est également un point important qui devra être traité, nous avons montré des techniques permettant d'y parvenir, dans un autre contexte, dans le chapitre précédent.

L'inférence bayésienne fait partie des thématiques de recherche très présentes au LSS en particulier dans le *Groupe Problèmes Inverses (GPI)*. J'envisage de mener à bien ce travail en profitant de l'expertise de mes collègues du GPI et en particulier au travers d'une collaboration avec Aurélia Fraysse, maître de conférences au GPI.

2. Le décodage itératif des turbo-codes que nous avons longuement discuté dans le chapitre 4 ainsi que l'algorithme de Gallager pour le décodage des codes LDPC sont des exemples d'implémentation de type *expectation propagation*.

Exposé détaillé

4 — Techniques itératives de décodage

OBJECTIF : Analyse, interprétation et amélioration du décodage itératif par des techniques issues de l'optimisation, de la théorie des jeux et à l'aide d'outils statistiques.

4.1 Introduction

La création des turbo-codes [32, 69] a marqué un tournant dans le domaine des communications en permettant d'atteindre des niveaux de performance inédits. Le décodage itératif, bien que sous-optimal permet d'obtenir des résultats en proximité des limites théoriques atteignables. Contrairement à bon nombre de systèmes modernes, l'algorithme mis en oeuvre n'a pas été obtenu comme solution d'un problème d'optimisation. Il est par conséquent difficile à analyser et les raisons de son efficacité sont longtemps demeurées mystérieuses. C'est un domaine de recherche qui a été très actif dans les années passées. Sans pouvoir être exhaustif, nous en donnons un aperçu ci-après. Les techniques de type EXIT-Chart [42] ou à évolution de densité [68, 73] sont des outils importants pour déterminer la convergence ou la stabilité d'un système donné. Elles sont particulièrement efficaces lorsque les blocs traités sont de taille relativement importante. Le lien entre les algorithmes de type *belief propagation* et le décodage itératif est étudié dans [132] mais la présence de boucles dans le graphe du turbo-code ne permet pas d'appliquer les résultats de convergence connus. Toujours dans le contexte des algorithmes de type *belief propagation*, de nombreuses contributions [144, 161] ont montré les liens entre le décodage turbo et les approximations de Bethe/Kikuchi utilisées en physique statistique. Il est en particulier prouvé que les points stationnaires du turbo-décodeur peuvent être obtenus à partir de l'énergie libre de Bethe. Ce résultat n'est pas limité aux seuls turbo-codes mais est une propriété générale des algorithmes de type *belief propagation* [172, 202]. Cette analogie ne permet pas d'aller beaucoup plus loin dans l'analyse de la convergence du système et ne donne pas non plus d'explication sur les très bonnes performances du décodage itératif. Elle permet en revanche de développer de nouvelles alternatives algorithmiques [23, 172]. Un autre ensemble de contributions [102, 143, 149, 171, 191] a considéré la géométrie de l'information comme outil pour étudier le décodage itératif. Ces travaux ont conduit à une interprétation élégante mais partielle de la procédure itérative en terme de projection sur des espaces de probabilité. L'écueil principal de cette approche est que la propagation d'extrinsèques (et non des probabilités *a posteriori*) ne s'interprète pas aisément à l'aide de projections.

C'est en suivant cette démarche que j'ai commencé à travailler sur ce sujet en 2007 ; c'est

également le point de départ de la thèse de Z. Naja [153]. Notre première contribution a été de montrer que la démodulation itérative dans un système BICM (*Bit Interleaved Coded Modulation*) est un algorithme de Dykstra [71] avec alternance entre *I-projection* et *rI-projection* [1]. L'algorithme de Dykstra diffère de l'algorithme des projections alternées par une étape supplémentaire qui permet de rendre compte de la propagation d'extrinsèques. Au delà d'une description appropriée du décodage itératif, la modélisation par un algorithme de Dykstra ne nous a pas permis d'établir de nouveaux résultats de convergence ni de proposer des améliorations à la procédure usuelle. Nous avons donc changé d'approche en nous inspirant des résultats de [192] dans lequel le décodage itératif est vu comme une solution itérative d'un problème d'optimisation sous contrainte dont la fonction d'objectif est la vraisemblance des séquences binaires. C'est une nouvelle interprétation très prometteuse puisqu'elle remet le décodage itératif dans le champs de la théorie de l'optimisation pour lequel de nombreux outils d'analyse (de convergence par exemple) sont disponibles. Ce sont les travaux effectués dans ce cadre là qui sont présentés dans ce chapitre. Nous allons détailler l'ensemble des étapes permettant d'obtenir le décodage itératif à partir de la vraisemblance sur les séquences binaires [2] et identifier les possibles degrés de liberté. Nous allons montrer comment la théorie des jeux permet de caractériser les points stationnaires de l'algorithme et nous établirons le lien entre la fonction d'utilité du jeu et la vraisemblance. Nous donnerons également un nouveau résultat de convergence et nous montrerons comment le décodeur est capable de s'auto-évaluer à l'aide de nouveaux critères de performance. Nous montrerons enfin comment calculer de manière optimale la "bonne quantité" d'extrinsèques à transmettre au fil des itérations.

4.2 Modèle et notations

Nous considérons un système de communication dans lequel un message binaire $\mathbf{b}$ de longueur n_b est codé puis transmis sur un canal sans mémoire. La séquence reçue est notée $\mathbf{r}_{obs}$. Sans *a priori* sur le message, l'estimation $\widehat{\mathbf{b}}$ de $\mathbf{b}$ à partir des observations peut s'écrire sous la forme suivante :

$$\widehat{\mathbf{b}} = \arg \max_{\mathbf{b}' \in \{0,1\}^{n_b}} p(\mathbf{r}_{obs}|\mathbf{b}') \tag{4.1}$$

où $p(\mathbf{r}_{obs}|\mathbf{b})$ est la fonction de vraisemblance qui dépend du système et du canal de transmission.

■ **Exemple 4.1** Un système de transmission à configuration série est représenté sur la figure 4.1. Il peut s'agir d'un système BICM (Bit Interleaved Coded Modulation). Dans ce cas l'émetteur est constitué d'un code convolutif et d'un *mapping* (association bit/symbole) séparés par un entrelaceur. Le même schéma de transmission s'applique aux turbo-codes à concaténation série en remplaçant le *mapping* par un second code convolutif. ■

FIGURE 4.1 – Système de transmission - Configuration série (SCTC ou BICM)

Dans ces deux exemples, la relation entre $\mathbf{b}$ et $\mathbf{d}$ est bijective, et le problème (4.1) peut s'écrire :

$$\widehat{\mathbf{d}} = \arg \max_{\mathbf{d}' \in \{0,1\}^{n}} p(\mathbf{r}_{obs}|\mathbf{d}') \tag{4.2}$$

1. Soit $D(p||q)$ la divergence de Kullback-Leibler entre p et q. Le minimum de $D(p||q)$ pour $p \in \mathscr{S}$ est noté $D(\mathscr{S},q)$. Si le minimum est unique il est appelé *I-projection* de q sur $\mathscr{S}$. De la même manière, le minimum de $D(p||q)$ pour $q \in \mathscr{T}$ est noté $D(p,\mathscr{T})$. Si le minimum est unique, il est appelé *rI-projection* de p sur $\mathscr{T}$ [58].

2. Contrairement à [192], nous ne considérons pas un problème d'optimisation avec contrainte. L'interprétation dans [192] n'est pas totalement satisfaisante car la contrainte n'est pas parfaitement définie, elle varie avec les itérations et est calculée *a posteriori* une fois la solution trouvée.

où n est la longueur de la séquence $\mathbf{d}$. Pour un turbo-code à concaténation parallèle, la formulation initiale doit être conservée. De manière générale, le problème d'optimisation se met sous la forme :

$$\widehat{\mathbf{x}} = \arg \max_{\mathbf{x}\in\{0,1\}^n} p(\mathbf{r}_{obs}|\mathbf{x}) \tag{4.3}$$

où $\mathbf{x}$ sera égal à $\mathbf{b}$ ou à $\mathbf{d}$ selon le système. Nous supposons dans ce chapitre que la vraisemblance $p(\mathbf{r}_{obs}|\mathbf{x})$ s'écrit sous la forme d'un produit $p(\mathbf{r}_{obs}|\mathbf{x}) = \widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x})$ qui reflète les deux étapes mises en œuvre à l'émetteur. Nous considérerons donc le problème d'optimisation sous la forme suivante :

$$\widehat{\mathbf{x}} = \arg \max_{\mathbf{x}\in\{0,1\}^n} \widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x}) \tag{4.4}$$

où $\widetilde{p}_1(\mathbf{x})$ et $\widetilde{p}_2(\mathbf{x})$ peuvent dépendre du vecteur d'observation.

■ **Exemple 4.2** Pour le BICM, $\widetilde{p}_1(\mathbf{x}) = p_{ch}(\mathbf{r}_{obs}|\mathbf{x})$ est la probabilité de recevoir la séquence $\mathbf{r}_{obs}$ pour une séquence codée $\mathbf{x}$, elle dépend du vecteur d'observation et du *mapping* mais pas du code. Le deuxième terme ne dépend que du code convolutif et s'écrit $\widetilde{p}_2(\mathbf{x}) = I_{cod}(\mathbf{x})$ où I_{cod} est la fonction indicatrice du code.

■

Ce problème est trop complexe pour une résolution directe. La section suivante est consacrée à l'obtention d'un critère sous-optimal à complexité réduite. Nous montrons l'ensemble des étapes qui conduisent de (4.4) au décodage itératif tel que nous le connaissons.

4.3 Décodage distribué

La complexité de (4.4) est due à la dimension du problème et à la difficulté à prendre en compte simultanément les contraintes ou la connaissance du système représentée par le produit $\widetilde{p}_1\widetilde{p}_2$. Pour lever ces deux verrous, on travaille usuellement sur les marginales $p_i(x_i)$, $1 \leq i \leq n$, et on considère séparément les deux composantes $\widetilde{p}_1$ et $\widetilde{p}_2$.

La première étape intermédiaire consiste à remplacer la variable d'optimisation $\mathbf{x}$, discrète, par une variable continue $p(\mathbf{x}) = \prod_{i=1}^{n} p_i(x_i)$. Le problème s'écrit alors :

$$\widehat{p}(\mathbf{x}) = \arg \max_{p(\mathbf{x})\in\mathscr{E}_S} \sum_{\mathbf{x}} \widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x}) \prod_{i=1}^{n} p_i(x_i) \tag{4.5}$$

où $\mathscr{E}_S$ est l'ensemble des lois de probabilité séparables. Il s'agit d'une écriture équivalente du problème (4.4) puisque le maximum est obtenu pour $\widehat{p}(\mathbf{x}) = \delta_{\mathbf{x},\widehat{\mathbf{x}}}$ [192], [16]. Nous écrirons dans la suite $p_i(x_i)$ sous la forme d'un produit $p_i(x_i) = p_{y,i}(x_i)p_{z,i}(x_i)$. On aura alors $p(\mathbf{x}) = \prod_{i=1}^{n} p_{y,i}(x_i)p_{z,i}(x_i)$. Nous verrons l'intérêt de cette écriture dans la section 4.4. En regroupant les termes de la somme relativement à la valeur de x_k, on peut écrire de manière tout à fait équivalente (4.5) sous la forme :

$$\widehat{p}(\mathbf{x}) = \arg \max_{p(\mathbf{x})\in\mathscr{E}_S} \sum_{x_k} \sum_{\mathbf{x}:x_k} \widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x}) \prod_{i=1}^{n} p_i(x_i) \tag{4.6}$$

La complexité calculatoire est encore élévée à cause de la prise en compte simultanée, dans le critère, de $\widetilde{p}_1(\mathbf{x})$ et $\widetilde{p}_2(\mathbf{x})$ par l'intermédiaire du produit $\widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x})$. Nous montrons dans ce chapitre que les équations de remise à jour usuelles du décodage itératif s'obtiennent à partir de n problèmes plus simples définis comme suit.

Définition 4.3.1 On définit une suite de n problèmes d'optimisation selon

$$\left(\widehat{p}_{y,k}\widehat{p}_{z,k}\right) = \arg \max_{p_{y,k}p_{z,k}\in\mathscr{F}} \widetilde{\mathscr{C}}_k(p_{y,k},p_{z,k}) \quad 1 \leq k \leq n \tag{4.7}$$

où $\mathscr{F}$ est l'ensemble des lois de probabilité sur $x_k \in \{0;1\}$ et avec

$$\widetilde{\mathscr{C}}_k(p_{y,k},p_{z,k}) \stackrel{def}{=} \sum_{x_k}\left(\sum_{\mathbf{x}:x_k}\widetilde{p}_1(\mathbf{x})\prod_i p_{z,i}(x_i)\right)\left(\sum_{\mathbf{x}':x_k}\widetilde{p}_2(\mathbf{x}')\prod_i p_{y,i}(x'_i)\right) \tag{4.8}$$

La différence majeure entre les problèmes (4.5) et (4.7) est visible au niveau du produit $\widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x}')$. Dans le problème initial, on ne considère que les produits pour lesquels $\mathbf{x} = \mathbf{x}'$ alors que dans $\widetilde{\mathscr{C}}_k$ on considère les termes $\mathbf{x}$ et $\mathbf{x}'$ tels que $x_k = x'_k$. Dans le critère $\widetilde{\mathscr{C}}_k$, la contribution du bit i, $i \neq k$ est évaluée séparément dans les deux composantes alors que la contribution du bit k est évaluée de manière conjointe. C'est pour cela que la valeur du bit k devra être estimée à partir de la maximisation du critère $\widetilde{\mathscr{C}}_k$. Dans le cas particulier ou $\widetilde{p}_1$ et $\widetilde{p}_2$ sont elles-mêmes des densités séparables, (4.7-4.8) est équivalent à (4.4). Dans le cas contraire, chaque critère $\widetilde{\mathscr{C}}_k$ est une approximation du critère optimal.

Notation 4.1. *Nous notons $\widetilde{\mathscr{C}}$ le critère approché moyen avec*

$$\widetilde{\mathscr{C}}(p_y,p_z) = \frac{1}{n}\sum_{k=1}^{n}\widetilde{\mathscr{C}}_k(p_{y,k},p_{z,k}) \tag{4.9}$$

On montre le résultat suivant.

Proposition 4.3.1 Si $\widetilde{\mathscr{C}}(p_y,p_z)$ a un maximum global et si ce maximum global est obtenu pour le couple $(\widehat{p}_y(\mathbf{x}),\widehat{p}_z(\mathbf{x}))$ tel que $\widehat{p}_y(\mathbf{x}) = \widehat{p}_z(\mathbf{x}) = \delta_{(\mathbf{x},\mathbf{x}_0)}$ alors $\mathbf{x}_0$ est solution de (4.3).

Démonstration. En organisant les termes impliqués dans $\widetilde{\mathscr{C}}(p_y,p_z)$ en fonction de la distance de Hamming $d_H(\mathbf{x},\mathbf{x}')$ entre $\mathbf{x}$ et $\mathbf{x}'$, on montre que si $\widetilde{\mathscr{C}}(p_y,p_z)$ a un maximum global en $\widehat{p}_y(\mathbf{x}) = \widehat{p}_z(\mathbf{x}) = \delta_{(\mathbf{x},\mathbf{x}_0)}$ alors $\widetilde{p}_1(\mathbf{x}_0)\widetilde{p}_2(\mathbf{x}_0) \geq (1-\frac{d_H(\mathbf{x},\mathbf{x}')}{n})\widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x}')\ \forall(\mathbf{x},\mathbf{x}')$ tel que $d_H(\mathbf{x},\mathbf{x}') \leq n-1$. En choisissant $\mathbf{x} = \mathbf{x}'$, on a $\widetilde{p}_1(\mathbf{x}_0)\widetilde{p}_2(\mathbf{x}_0) \geq \widetilde{p}_1(\mathbf{x})\widetilde{p}_2(\mathbf{x})$. ■

Nous verrons par la suite que p_y et p_z sont les probabilités extrinsèques propagées au cours du décodage itératif alors que le produit $p_y p_z$ est (à une constante de normalisation près) la probabilité *a posteriori* (APP). D'après la proposition 4.3.1, si l'APP obtenue après maximisation du critère est une mesure de Dirac alors cette mesure nous donne la valeur de l'optimum du problème (4.4). Nous étudions dans la suite le processus itératif permettant de résoudre le problème approché (4.7-4.8).

4.4 Procédure itérative d'optimisation

Nous considérons ici la stratégie de maximisation distribuée (4.7-4.8). Les marginales $p_{y,k}$, $p_{z,k}$ sont estimées à partir du critère $\widetilde{\mathscr{C}}_k$ dans lequel les autres marginales sont supposées fixes.

Notation 4.2. *Nous utiliserons dans la suite la notation $f_{x_k}(u_{[k]},\widetilde{p}_\alpha) = \sum_{\mathbf{x}:x_k}\widetilde{p}_\alpha(\mathbf{x})\prod_{j\neq k}u_j(x_j)$ où $u_{[k]} = (u_1,u_2,...,u_{k-1},u_{k+1},...,u_n)$. Nous utiliserons également la notation $\overline{x}_k$ telle que $\overline{x}_k = 1-x_k$ avec $x_k \in \{0;1\}$.*

Le problème d'optimisation se réécrit alors sous la forme :

$$\left(\widehat{p}_{y,k}\widehat{p}_{z,k}\right) = \arg \max_{p_{y,k}p_{z,k}\in\mathscr{F}} \sum_{x_k} p_{y,k}(x_k)p_{z,k}(x_k)f_{x_k}(p_{z,[k]},\widetilde{p}_1)f_{x_k}(p_{y,[k]},\widetilde{p}_2) \tag{4.10}$$

C'est un critère linéaire par rapport à la variable d'optimisation. La solution est donc située sur les bords de l'intervalle. La forme binaire des solutions est de nature à bloquer le processus itératif dans un minimum local. Pour limiter cet effet, on peut remplacer les décisions "dures" sur chacun des bits par des décisions souples de la forme :

$$\widehat{p}_{y,k}(x_k)\widehat{p}_{z,k}(x_k) \propto \left(f_{x_k}(p_{z,[k]}, \widetilde{p}_1) f_{x_k}(p_{y,[k]}, \widetilde{p}_2) \right)^{\alpha} \tag{4.11}$$

où α est une constante positive. Après une étape de normalisation, on peut voir que la loi ci-dessus tend vers la solution de (4.10) lorsque $\alpha \to \infty$ et tend vers une loi uniforme lorsque $\alpha \to 0$. Ce paramètre est en quelque sorte un indice de la confiance accordée à la solution trouvée. On peut donc penser que la valeur de α devrait être relativement faible au début du processus itératif et augmenter au fil des itérations. Nous nous intéresserons aux valeurs à donner à α plus loin dans ce chapitre.

Le passage à des décisions souples peut être compris comme l'ajout d'un terme de régularisation au critère (4.10). En effet (4.11) est solution du problème régularisé :

$$\left(\widehat{p}_{y,k}\widehat{p}_{z,k} \right) = \arg \max_{p_{y,k}p_{z,k}} \alpha \sum_{x_k} p_{y,k}(x_k) p_{z,k}(x_k) \log \left(\frac{f_{x_k}(p_{z,[k]}, \widetilde{p}_1) f_{x_k}(p_{y,[k]}, \widetilde{p}_2)}{\sum_{x_k} f_{x_k}(p_{z,[k]}, \widetilde{p}_1) f_{x_k}(p_{y,[k]}, \widetilde{p}_2)} \right) + H(p_{y,k}p_{z,k}) \tag{4.12}$$

où $H(p_{y,k}p_{z,k}) = -\sum_{x_k} p_{y,k}(x_k)p_{z,k}(x_k) \log \left(p_{y,k}(x_k)p_{z,k}(x_k) \right) + p_{y,k}(x_k)p_{z,k}(x_k)$ est la définition de l'entropie [27] pour $p_{y,k}(x_k)p_{z,k}(x_k) \in [0; +\infty[$. La métrique dans (4.12) prend la forme d'une énergie libre, on reconnaît l'expression d'une énergie variationelle moyenne dans le premier terme et l'expression d'une entropie dans le second [172, 202]. La solution (4.12) caractérise le produit $p_{y,k}p_{z,k}$, les expressions de $p_{y,k}$ et $p_{z,k}$ dépendent de l'ordonnancement. Nous définissons ci-dessous le choix qui conduit à l'expression des probabilités extrinsèques usuelles dans les turbo-codes. Il s'agit d'une implémentation de type Jacobi/Gauss-Seidel.

Résultat 4.4.1 En suivant l'ordonnancement ci-dessous :

— *Initialisation :* $p_{y,k}^{(0)}(x_k) = p_{z,k}^{(0)}(x_k) = \frac{1}{2}$, $1 \leq k \leq n$ et $x_k \in \{0,1\}$.

— *Itération it (Etape 1)* : Calcul de $p_{z,k}^{(it)}$ connaissant $p_{y,k}^{(it-1)}$ et $p_{z,i}^{(it-1)}$, $i \neq k$ et $1 \leq k \leq n$ selon

$$p_{y,k}^{(it-1)}(x_k) p_{z,k}^{(it)}(x_k) \propto \left(f_{x_k}(p_{z,[k]}^{(it-1)}, \widetilde{p}_1) f_{x_k}(p_{y,[k]}^{(it-1)}, \widetilde{p}_2) \right)^{\alpha}$$

— *Itération it (Etape 2)* : Calcul de $p_{y,k}^{(it)}$ connaissant $p_{z,k}^{(it)}$ et $p_{y,i}^{(it-1)}$, $i \neq k$ et $1 \leq k \leq n$ selon

$$p_{y,k}^{(it)}(x_k) p_{z,k}^{(it)}(x_k) \propto \left(f_{x_k}(p_{z,[k]}^{(it)}, \widetilde{p}_1) f_{x_k}(p_{y,[k]}^{(it-1)}, \widetilde{p}_2) \right)^{\alpha}$$

On obtient

$$\ell_{z,k}^{(it)} \stackrel{\text{def}}{=} \log \left(\frac{p_{z,k}^{(it)}(x_k)}{p_{z,k}^{(it)}(\overline{x_k})} \right) = \alpha \log \left(\frac{\sum_{\mathbf{x}:x_k} \widetilde{p}_2(\mathbf{x}) \prod_{j \neq k} p_{y,j}^{(it-1)}(x_j)}{\sum_{\mathbf{x}:\overline{x_k}} \widetilde{p}_2(\mathbf{x}) \prod_{j \neq k} p_{y,j}^{(it-1)}(x_j)} \right)$$

$$\ell_{y,k}^{(it)} \stackrel{\text{def}}{=} \log \left(\frac{p_{y,k}^{(it)}(x_k)}{p_{y,k}^{(it)}(\overline{x_k})} \right) = \alpha \log \left(\frac{\sum_{\mathbf{x}:x_k} \widetilde{p}_1(\mathbf{x}) \prod_{j \neq k} p_{z,j}^{(it)}(x_j)}{\sum_{\mathbf{x}:\overline{x_k}} \widetilde{p}_1(\mathbf{x}) \prod_{j \neq k} p_{z,j}^{(it)}(x_j)} \right)$$

Les expressions des extrinsèques, exprimées sous la forme de logarithme du rapport de vraisemblance (LLR), fournies par un algorithme de type BCJR correspondent aux expressions de $\ell_{y,k}$ et

$\ell_{z,k}$ ci-dessus avec $\alpha = 1$. Le décodage itératif peut donc être vu comme une procédure distribuée dont l'objectif est de rendre maximum le critère $\widetilde{\mathscr{C}}$. Nous avons toutefois vu que le lien entre l'expression des LLR et le critère $\widetilde{\mathscr{C}}$ n'est pas direct. Il passe par le découpage de $\widetilde{\mathscr{C}}$ en une somme de n fonctions $\widetilde{\mathscr{C}_k}$ pour lesquelles l'optimisation est conduite sur une jeu limité de variables après régularisation du critère. La théorie des jeux est un cadre théorique qui permet d'analyser et de décrire plus simplement cette procédure. C'est l'objet de la section suivante.

4.5 Application de la théorie des jeux au décodage itératif

Les notions de théorie des jeux utilisées ici se limitent aux définitions de base. De nombreux exemples d'application de la théorie des jeux au traitement du signal et aux télécommunications sont donnés dans [120] ou encore dans [121] auxquels le lecteur peut se référer pour avoir une vue plus large que celle donnée ici. Nous définissons dans un premier temps la notion de jeu [159] telle que nous l'entendons dans ce document.

Définition 4.5.1 — Jeu. Un jeu $\mathscr{G}$ est défini par un triplet $\mathscr{G} = (\mathscr{N}, \{S_i\}_{i\in\mathscr{N}}, \{u_i\}_{i\in\mathscr{N}})$ où $\mathscr{N} = \{1,2,...,n\}$ est un ensemble fini de joueurs, $\forall i \in \mathscr{N}$, $\{S_i\}$ est l'ensemble des stratégies possibles pour le joueur i et u_i sa fonction d'utilité.

■ **Exemple 4.3** Nous considérons dans ce chapitre les deux jeux suivants :

— *Jeu 1 (solutions souples)*. On note $\mathscr{G}_S$ un jeu à n joueurs dans lequel le joueur i cherche à maximiser sa fonction d'utilité u_i^S définie par

$$u_i^S(\ell_{y,i},\ell_{z,i},\ell_{y,[i]},\ell_{z,[i]}) = -\left\| \ell_{y,i} + \ell_{z,i} - \alpha \log\left(\frac{f_{x_i}(p_{z,[i]},\widetilde{p}_1) f_{x_i}(p_{y,[i]},\widetilde{p}_2)}{f_{\overline{x_i}}(p_{z,[i]},\widetilde{p}_1) f_{\overline{x_i}}(p_{y,[i]},\widetilde{p}_2)} \right) \right\|^2 \tag{4.13}$$

Le jeu de variables affecté au joueur i est $(\ell_{y,i},\ell_{z,i})$ alors que le jeu de variables des autres joueurs est $(\ell_{y,[i]},\ell_{z,[i]})$. L'ensemble des stratégies du joueur i est $\{S_i^S\} = \mathbb{R}^2$.

— *Jeu 2 (solutions binaires)*. On note $\mathscr{G}_B$ un jeu à n joueurs dans lequel le joueur i cherche à maximiser sa fonction d'utilité $u_i^B = \widetilde{\mathscr{C}_i}$ définie par l'équation (4.8). Le jeu de variables affecté au joueur i est $(p_{y,i},p_{z,i})$ alors que le jeu de variables des autres joueurs est $(p_{y,[i]},p_{z,[i]})$. L'ensemble des stratégies du joueur i est $\{S_i^B\} = [0,1]^4$.

■

Dans chacun des deux jeux $\mathscr{G}_S$ et $\mathscr{G}_B$, les joueurs sont les éléments binaires constituant le message et pas les deux éléments constitutifs de la chaîne de mise en forme (codeur 1/codeur 2 ou *mapping*/code) comme on aurait pu le penser *a priori*. Nous donnons maintenant la définition d'un équilibre de Nash.

Définition 4.5.2 — Equilibre de Nash. Un état $\mathbf{s}^*$ est un équilibre de Nash (NE) du jeu $\mathscr{G}$ si

$$\forall i \in \mathscr{N}, \forall s_i' \in S_i,\ \ u_i(s_i^*, s_{[i]}^*) \geq u_i(s_i', s_{[i]}^*)$$

Si l'égalité est vraie au sens strict pour tous les joueurs, on parlera alors d'équilibre de Nash strict.

Nous considérerons ici uniquement les équilibres purs dans lesquels les stratégies des joueurs consistent à choisir un élément dans l'ensemble $\{S_i\}$ de manière déterministe. Nous ne considérerons pas les stratégies mixtes dans lesquelles une probabilité est associée à chaque valeur de $\{S_i\}$, le joueur jouant alors en fonction de la loterie correspondante. Un équilibre de Nash est une situation dans laquelle chaque joueur a trouvé la meilleure stratégie compte-tenu des stratégies des autres joueurs. Le jeu est stable dans la mesure ou si un joueur décide seul de modifier sa stratégie, il s'affaiblira. L'existence d'un équilibre de Nash n'implique pas l'optimalité, il peut exister des

choix différents des joueurs qui conduisent à des valeurs supérieures des fonctions d'utilité. En relation avec les jeux, $\mathscr{G}^S$ et $\mathscr{G}^B$ nous montrons les résultats suivants :

Proposition 4.5.1 Soit (ℓ_y^*, ℓ_z^*) un couple de séquences de LLR ayant pour élément i le couple $(\ell_{y,i}^*, \ell_{z,i}^*)$ tel que :

$$\ell_{y,i}^* + \ell_{z,i}^* = \alpha \log\left(\frac{f_{x_i}(p_{z,[i]}^*, \widetilde{p}_1) f_{x_i}(p_{y,[i]}^*, \widetilde{p}_2)}{f_{\overline{x_i}}(p_{z,[i]}^*, \widetilde{p}_1) f_{\overline{x_i}}(p_{y,[i]}^*, \widetilde{p}_2)}\right) \quad \forall i \in \{0,1,...,n\} \tag{4.14}$$

Alors, (ℓ_y^*, ℓ_z^*) est un équilibre de Nash du jeu $\mathscr{G}^S$.

Démonstration. Si (ℓ_y^*, ℓ_z^*) est solution de (4.14) alors $(\ell_{y,i}^*, \ell_{z,i}^*)$ est un maximum global de $u_i^S(\ell_{y,i}, \ell_{z,i}, \ell_{y,[i]}^*, \ell_{z,[i]}^*)$. Comme cela est vrai $\forall i \in \{0,1,...,n\}$, (ℓ_y^*, ℓ_z^*) est un NE de $\mathscr{G}^S$. Si la solution de (4.14) est unique alors (ℓ_y^*, ℓ_z^*) est l'unique maximum global de u_i^S pour $1 \leq i \leq n$ et (ℓ_y^*, ℓ_z^*) est un NE strict. ■

Le lien entre le point fixe (ℓ_y^*, ℓ_z^*) et le jeu $\mathscr{G}^B$ est établi dans la proposition suivante.

Proposition 4.5.2 Soit (ℓ_y^*, ℓ_z^*) un couple de séquences de LLR dont l'élément $(\ell_{y,i}^*, \ell_{z,i}^*)$ satisfait (4.14) pour tout $i \in \{0,1,...,n\}$. Alors $\mathscr{B}(p_y^*, p_z^*)$ est un point d'équilibre *induit* du jeu $\mathscr{G}^B$; c'est à dire :

$$\forall i \in \{1,...,n\}, \forall s_i = (p_{y,i}(x_i), p_{z,i}(x_i)) \in S_i = [0;1]^4, \;\; \widetilde{\mathscr{C}_i}(\mathscr{B}(s_i^*), s_{-i}^*) \geq \widetilde{\mathscr{C}_i}(s_i, s_{-i}^*)$$

où $\mathscr{B}()$ est une fonction de $[0,1]^4$ vers $\{(0,0);(1,1)\}$ qui associe au couple $s_i^* = (p_{y,i}^*(x_i), p_{z,i}^*(x_i))$ le couple (x_i, x_i) si $p_{y,i}^*(x_i) p_{z,i}^*(x_i) > p_{y,i}^*(\overline{x_i}) p_{z,i}^*(\overline{x_i})$ et le couple $(\overline{x_i}, \overline{x_i})$ sinon.

Démonstration. Si $({\ell_y}^*, \ell_z^*)$ est solution de (4.14) alors

$$p_{y,i}^*(x_i) p_{z,i}^*(x_i) \propto \left(f_{x_i}({p_{z,[i]}}^*, \widetilde{p}_1) f_{x_i}(p_{y,[i]}^*, \widetilde{p}_2)\right)^{\alpha}$$

et $\mathscr{B}(p_{y,i}^*(x_i), p_{z,i}^*(x_i))$ est solution du problème d'optimisation (4.10). ■

L'équilibre que nous venons de définir ne peut être qualifié d'équilibre de Nash ; il diffère de la définition 4.5.2 par l'introduction de la fonction $\mathscr{B}$. Cette fonction $\mathscr{B}$ correspond à la prise de décision, binaire, sur la base de la probabilité *a posteriori* $p_{y,i} p_{z,i}$. C'est tout de même un équilibre puisque un choix unilatéral du joueur i autre que $\mathscr{B}(p_{y,i}^*(x_i), p_{z,i}^*(x_i))$ conduirait à une valeur plus faible de la fonction d'utilité $\widetilde{\mathscr{C}_i}$. En d'autres termes, une fois qu'un point d'équilibre (4.14) a été atteint, la meilleure décision à prendre pour chacun des joueurs est la décision dure (binaire) déduite de la valeur de la probabilité *a posteriori* $p_y p_z$ ou de manière équivalente de la LLR $\ell_y + \ell_z$.

Définition 4.5.3 — Utilité Moyenne. On appelle utilité moyenne associée au jeu $\mathscr{G}$, la fonction W définie selon

$$W = \frac{1}{n} \sum_{i=1}^{n} u_i$$

où u_i est la fonction d'utilité du joueur i.

Cette fonction W est souvent appelée bien-être social (*social welfare*) [22] du jeu dans la littérature et est utilisée comme une mesure d'efficacité d'une société. Plus prosaïquement ici, nous nous

servirons de cette mesure pour évaluer l'efficacité du processus d'optimisation. Nous considérons à nouveau les deux jeux $\mathscr{G}^S$ et $\mathscr{G}^B$.

■ **Exemple 4.4** L'utilité moyenne du jeu $\mathscr{G}^S$ est

$$W^S(\ell_y,\ell_z)=\frac{1}{n}\sum_{i=1}^{n}u_i^S(\ell_{y,i},\ell_{z,i},\ell_{y,[i]},\ell_{z,[i]}) \tag{4.15}$$

Si (ℓ_y^*,ℓ_z^*) est une équilibre de Nash du jeu $\mathscr{G}^S$, on a $W^S(\ell_y^*,\ell_z^*)=0$. Cette mesure n'amène donc aucune information sur la proximité du point d'équilibre et de la solution. ■

■ **Exemple 4.5** L'utilité moyenne du jeu $\mathscr{G}^B$ est

$$W^B(p_y,p_z)=\frac{1}{n}\sum_{i=1}^{n}u_i^B(p_{y,i},p_{z,i},p_{y,[i]},p_{z,[i]})=\widetilde{\mathscr{C}} \tag{4.16}$$

où $\widetilde{\mathscr{C}}$ est le critère approché moyen défini en (4.9) et déduit du critère maximum de vraisemblance. Nous pourrons donc utiliser W^B comme un critère de performance. Cette mesure fait le lien entre le critère optimal au sens du maximum de vraisemblance et la procédure itérative. ■

Chaque fonction d'utilité $u_i=\widetilde{\mathscr{C}_i}$ du jeu $\mathscr{G}^B$ est à valeur dans l'intervalle $[0;1]$. La valeur maximale de $\widetilde{\mathscr{C}}$ est donc 1. Si nous obtenons cette valeur pour $\widetilde{\mathscr{C}}$ à l'issue du décodage itératif, nous avons trouvé un maximum global du problème. On peut montrer que cette valeur n'est atteignable que si p_yp_z est une mesure de Dirac. Nous avons vu, dans la proposition 4.3.1, que dans ce cas la solution obtenue est le maximum de vraisemblance. Nous cherchons donc des solutions pour lesquelles $\widetilde{\mathscr{C}}$ est au voisinage de 1, la solution trouvée sera alors le maximum de vraisemblance.

Pour le moment, nous n'avons pas prouvé l'existence d'un ou de plusieurs équilibres de Nash. Nous nous intéressons maintenant à cette question. On prouve l'existence d'un équilibre de Nash en prouvant l'existence d'une solution dans un problème de point fixe. Pour le décodage itératif, le problème du point fixe à considérer est donné ci-dessous :

$$\ell_{y,k} = \alpha\log\left(\frac{f_{x_k}(p_{z,[k]},\widetilde{p}_1)}{f_{\bar{x}_k}(p_{z,[k]},\widetilde{p}_1)}\right) \quad 1\le k\le n \tag{4.17}$$

$$\ell_{z,k} = \alpha\log\left(\frac{f_{x_k}(p_{y,[k]},\widetilde{p}_2)}{f_{\bar{x}_k}(p_{y,[k]},\widetilde{p}_2)}\right) \quad 1\le k\le n \tag{4.18}$$

Ce sont les deux équations qui sont satisfaites si la procédure itérative converge. Nous écrivons ce système d'équations sous la forme équivalente :

$$\ell_{y,k}+\ell_{z,k} = \alpha[\Pi_{\widetilde{p}_1}(\ell_z)]_k-(\alpha-1)\ell_{z,k} \quad 1\le k\le n \tag{4.19}$$

$$\ell_{y,k}+\ell_{z,k} = \alpha[\Pi_{\widetilde{p}_2}(\ell_y)]_k-(\alpha-1)\ell_{y,k} \quad 1\le k\le n \tag{4.20}$$

où $\Pi_{\widetilde{p}_1}(\ell_z)$ et $\Pi_{\widetilde{p}_2}(\ell_y)$ sont deux vecteurs de LLR avec pour élément $[\Pi_{\widetilde{p}_1}(\ell_z)]_k=\log\left(\frac{f_{x_k}(p_{z,[k]},\widetilde{p}_1)}{f_{\bar{x}_k}(p_{z,[k]},\widetilde{p}_1)}\right)+\ell_{z,k}$ et $[\Pi_{\widetilde{p}_2}(\ell_y)]_k=\log\left(\frac{f_{x_k}(p_{y,[k]},\widetilde{p}_2)}{f_{\bar{x}_k}(p_{y,[k]},\widetilde{p}_2)}\right)+\ell_{y,k}$.

Notation 4.3. *Nous utiliserons également la notation abrégée* $F_{\widetilde{p},\alpha}(\ell_u)=\alpha\Pi_{\widetilde{p}}(\ell_u)-(\alpha-1)\ell_u$.

L'existence de points fixes, dans le cas particulier, $\alpha=1$ a été montrée dans [171]. Nous généralisons ce résultat aux valeurs de α dans l'intervalle $]0;+\infty[$.

Proposition 4.5.3 Pour toute loi de probabilité $\widetilde{p}$, l'application $F_{\widetilde{p},\alpha}$ est un homéomorphisme $\forall \alpha \leq \alpha_{NE}$ où $\alpha_{NE} \geq 1$.

La démonstration de cette proposition est donnée dans [17] en annexe de ce document. L'application $F_{\widetilde{p},\alpha}$ caractérise l'effet d'une composante du récepteur (décodeur 1 ou décodeur 2 ou opération de *demapping*). Elle fait le lien entre l'*a priori* fourni à l'entrée et l'*a posteriori* obtenu en sortie. Montrer que $F_{\widetilde{p},\alpha}$ est un homéomorphisme revient à dire que l'on pourrait "théoriquement" construire le décodage itératif à l'envers en considérant que l'entrée de chaque composante est l'*a posteriori* $\ell_y + \ell_z$ et que la sortie obtenue est l'*a priori* ℓ_y ou ℓ_z. A partir de ce résultat, on peut montrer l'existence de points fixes dans le cas général.

Théorème 4.5.4 Il existe toujours $\alpha_{NE} \geq 1$ tel que $\forall \alpha \leq \alpha_{NE}$, le système (4.19-4.20) possède une solution.

La preuve est identique à celle donnée dans [171] en remplaçant $F_{\widetilde{p},1}$ par $F_{\widetilde{p},\alpha}$. On en déduit le résultat suivant concernant l'existence d'équilibres de Nash.

Corollaire 4.5.5 Le jeu $\mathscr{G}^S$ possède toujours au moins un NE tant que $\alpha \leq \alpha_{NE}$. Le jeu $\mathscr{G}^B$ possède toujours au moins un NE induit tant que $\alpha \leq \alpha_{NE}$.

Nous avons prouvé l'existence d'au moins un point fixe pour des valeurs de α dans l'intervalle $[0;\alpha_{NE}]$ où α_{NE} est propre au système. Nous nous intéressons maintenant à la convergence du processus itératif. Nous déterminons les conditions à vérifier pour garantir la convergence vers l'un des équilibres de Nash du problème.

Notation 4.4. *Nous utilisons la notation compacte $S_\alpha \ell_{yz} = 0$ pour le système d'équations (4.19-4.20) avec $\ell_{yz} = [\ell_y \;\; \ell_z]^T$.*

Nous utilisons également la notation $\mathbf{C}_a$, $a \in \{1,2\}$ qui représente une matrice dont les éléments sont donnés par

$$[\mathbf{C}_a]_{i,j} = p_a[x_j = 1 | x_i = 1] - p_a[x_j = 1 | x_i = 0]$$

où $p_1(\mathbf{x}) = K_1 \widetilde{p}_1(\mathbf{x}) p_z(\mathbf{x})$ et $p_2 = K_2 \widetilde{p}_2(\mathbf{x}) p_y(\mathbf{x})$ et où K_1, K_2 sont les constantes de normalisation associées.

Le jacobien ∇S_α de S_α est :

$$\nabla S_\alpha = \begin{pmatrix} \mathbf{I} & \alpha(\mathbf{I} - \mathbf{C}_1) \\ \alpha(\mathbf{I} - \mathbf{C}_2) & \mathbf{I} \end{pmatrix} \tag{4.21}$$

où $\mathbf{I}$ est la matrice identité de taille $n \times n$. Ce jacobien a été utilisé dans [192] (dans le cas particulier $\alpha = 1$) dans le but d'étudier la convergence du décodage itératif. Il y est montré que si ∇S_α est une M-matrice [3] alors le décodeur turbo converge vers une solution unique quel que soit le point d'initialisation. Vérifier que ∇S_α est une M-matrice n'est pas simple à réaliser en pratique. Nous proposons ici un point de vue différent. Au lieu de chercher à déterminer les conditions de convergence dans le cas $\alpha = 1$, nous montrons qu'il existe toujours des valeurs de α pour lesquelles la convergence de l'algorithme vers un point d'équilibre est garantie.

Théorème 4.5.6 Soit $S_\alpha : D \subset \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ une application continue sur l'ensemble D, et supposons que $S_\alpha \ell_{yz} = 0$ a une solution $\ell^*_{yz} \in D$, et qu'il existe $r \geq 0$, $Q = \{\ell_{yz} \in \mathbb{R}^{2n} : \| \ell_{yz} - \ell^*_{yz} \|_\infty \geq r\} \subset D$. Alors, il existe toujours α_0 tel que $\forall \alpha \leq \alpha_0$, ∇S_α est une matrice à diagonale dominante

3. Une M-matrice est une matrice carré ayant des éléments non-positif hors diagonale et des valeurs propres réelles et positives. Des formulations différentes mais équivalentes sont données dans [97].

stricte, ℓ^*_{yz} est unique dans Q, et pour tout $\ell^{(0)}_{yz}$ de Q les séquences de Jacobi et de Gauss-Seidel sont bien définies et convergent vers ℓ^*_{yz}.

Démonstration. — La continuité de S_α a été montrée dans la proposition 4.5.3.
— L'existence de solutions de $\mathbb{R}^{2n}$ au problème $S_\alpha \ell_{yz} = 0$ a été démontrée pour $\alpha \leq \alpha_{NE}$.
— Soit $\ell^{(k)}_{yz,i}$ la valeur de $\ell_{yz,i}$ à l'itération k, si $|\ell^{(k)}_{yz,i}| < +\infty$ alors Q existe.
— Nous avons montré dans la proposition 3 de [13] qu'il existe toujours $\alpha_C > 0$ tel que, $\forall \alpha \leq \alpha_C$, ∇S_α est une matrice à diagonale dominante stricte pour tout $\ell_{yz} \in \mathbb{R}^{2n}$.

Il suffit alors de choisir $\alpha_0 = min(\alpha_{NE}, \alpha_C)$ puis d'appliquer le théorème 5.7 de [146] pour prouver la convergence vers un point fixe. ■

La convergence est donc toujours possible en choisissant une valeur de α adaptée, éventuellement faible. La convergence est alors garantie vers un minimum local mais pas nécessairement vers l'optimum global.
Les résultats obtenus à ce stade sont les suivants :

— nous avons montré le lien entre le maximum de vraisemblance et le décodage itératif et nous avons identifié les approximations mises en jeux,
— nous avons montré pourquoi on devait propager des extrinsèques et pas des probabilité *a posteriori*,
— nous avons donné une interprétation du décodage itératif à l'aide de la théorie des jeux,
— grâce à la théorie des jeux, nous avons montré que les points fixes de l'algorithme de décodage sont des équilibres (induits) de Nash d'un jeu dont les fonctions d'utilité sont chacune une approximation du maximum de vraisemblance,
— nous avons analysé le processus itératif : sa convergence, ses points fixes,
— nous avons introduit un critère $\widetilde{\mathscr{C}}$ (utilité moyenne du jeu) qui est une approximation du maximum de vraisemblance et qui permet de juger de la proximité entre la solution trouvée et la solution optimale (au sens du maximum de vraisemblance) et qui peut être utilisé comme détecteur d'erreur ou comme critère d'arrêt au niveau du récepteur puisqu'il est calculé à partir des seules extrinsèques,
— nous avons montré que le maximum global de $\widetilde{\mathscr{C}}$ coïncide avec le maximum de vraisemblance,
— nous avons introduit un paramètre α qui, en fonction de la valeur choisie, est un compromis entre une absence de confiance sur la probabilité *a posteriori* ($\alpha = 0$, loi uniforme) ou une confiance totale ($\alpha \to \infty$, mesure de Dirac).

A ce stade, nous n'avons pas donné de moyen pratique pour régler correctement la valeur de α. De nombreux travaux ont été menés dans ce domaine et sur différents systèmes : BICM, turbo-codes ou codes LDPC en particulier lorsque le décodage est réalisé avec l'algorithme *min-sum* [75, 193]. Parmi les travaux récents sur ce thème nous pouvons citer [49], [19], [111], [204]. Les solutions proposées sont en général restreintes à une structure particulière. Bon nombre d'entre elles sont non-adaptatives ou nécessitent la connaissance ou une estimation acceptable du message transmis. On pourra trouver dans [25] de nombreuses références à des travaux portant sur le choix et/ou l'influence du facteur de pondération α sur les performances. Il est également montré dans [25] qu'un choix optimal du facteur de pondération conduit à une amélioration des performances pour un turbo-code avec un décodage MAP ou MaxLogMAP et également pour un turbo-code double-binaire avec décodage MaxLogMAP.

Dans la deuxième partie de ce chapitre, nous allons proposer une méthode générale, adaptative et basée sur des quantités présentes au niveau du récepteur pour évaluer la valeur à donner au paramètre α. Ce qui nous distingue de nombreuses publications antérieures est que la méthode proposée ne nécessite pas une connaissance explicite du rapport signal à bruit ou de l'itération, elle ne s'appuie pas non plus sur une étape d'apprentissage pour l'obtention du facteur de pondération optimal. Enfin, la méthodologie proposée est applicable à divers systèmes et à divers choix d'algorithmes

de décodage. Dans cette deuxième partie de chapitre, nous ne parlerons plus d'optimisation ou de théorie des jeux. Nous considérons à la place une approche statistique et nous revisitons la notion d'Exit Chart afin de simuler le comportement du système dans sa globalité.

4.6 Information mutuelle entre extrinsèques

4.6.1 Modélisation et hypothèses

Notation 4.5. *Nous noterons les variables aléatoires (v.a) à l'aide de lettres majuscules et leurs réalisations à l'aide de lettres minuscules. En particulier :*

- *X est la v.a. associée au message binaire avec $X \in \{-1;+1\}$,*
- *L_y et L_z sont les v.a. associées aux log-ratio des probabilités extrinsèques. Nous utiliserons également la notation L lorsqu'un résultat ou une propriété s'applique indifféremment à L_y ou à L_z.*

Les deux variables L_y et L_z sont des estimateurs de X. Le modèle communément utilisé pour relier L à X [42] est

$$L = \frac{\sigma^2}{2}X + \sigma N \tag{4.22}$$

où N est une v.a. gaussienne de loi $\mathcal{N}(0,1)$ et $\sigma \in \mathbb{R}^+$.

Notation 4.6. *On note $G_1(X)$ l'ensemble des v.a. définies selon le modèle (4.22).*

La pertinence de ce modèle a été testée dans [76] pour des turbo-codes et dans [77] pour des codes LDPC. On peut également noter que la pertinence du modèle gaussien ainsi que les propriétés de symétrie ont été validées dans [178] pour un turbo-détecteur constitué d'un égaliseur et d'un décodeur MAP. Pour les turbo-codes et les LDPC, il ressort de [76, 77] que l'approximation par une gaussienne est pertinente à condition de considérer la moyenne et la variance comme deux quantités indépendantes. Nous considérerons donc dorénavant le modèle ci-après :

$$L = \frac{\alpha\sigma^2}{2}X + \sigma N \tag{4.23}$$

avec $\alpha \in \mathbb{R}^{+*}$.

Notation 4.7. *On note $G_\alpha(X)$ l'ensemble des v.a. définies selon le modèle (4.23) où α est un réel positif fixé.*

On peut vérifier sans difficulté que la distribution de la variable $L|X$ vérifie la propriété suivante.

Propriété 4.6.1 Si L suit le modèle (4.23) alors $p_L(\ell|X) = p_L(-\ell|-X)$ et $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\alpha\ell}$.

La dernière relation fait le lien avec les sections précédentes et montre que la quantité à propager au cours des itérations est αL où α doit être déterminé. L'expression de l'information mutuelle $I(L;X)$ est donnée ci-dessous pour chacun des modèles considérés.

■ **Exemple 4.6** L suit le modèle (4.22), dans ce cas

$$I(L,X) = J(\sigma) = 1 - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\ell - \frac{\sigma^2}{2}\right)^2}{2\sigma^2}} \log_2(1+e^{-\ell})d\ell \tag{4.24}$$

où J est une fonction strictement croissante de la variable σ. ■

■ **Exemple 4.7** L suit le modèle (4.23), dans ce cas

$$\begin{aligned} I(L,X) &= 1-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\alpha\sigma}e^{-\frac{\left(\ell-\alpha\mu\right)^2}{2\alpha^2\sigma^2}}\log_2(1+e^{-\ell})d\ell \\ &= 1-\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\alpha\sigma}e^{-\frac{\left(\ell-\frac{\alpha^2\sigma^2}{2}\right)^2}{2\alpha^2\sigma^2}}\log_2(1+e^{-\ell})d\ell \\ &= J(\alpha\sigma) \end{aligned} \tag{4.25}$$

Il est intéressant de noter que, même si dans le modèle (4.23), L dépend de deux paramètres, $I(L;X)$ reste une fonction strictement croissante d'une seule variable. ■

Nous pouvons maintenant préciser le lien entre les deux modèle par le biais de l'information mutuelle.

Résultat 4.6.1 Soit $L \in G_\alpha(X)$, on a $I(L;X) = I(\alpha L;X)$ et $\alpha L \in G_1(X)$.

La multiplication de L par un scalaire ne change donc pas son information mutuelle. On peut toujours à partir de $L \in G_\alpha(X)$ obtenir l'élément de $G_1(X)$ ayant la même information mutuelle avec X : il suffit pour cela de multiplier L par α. Dans un EXIT-Chart, on évalue le comportement d'un élément de la chaîne de décodage en traçant l'évolution de l'information mutuelle $I(L;X)$ en sortie en fonction de l'information mutuelle en entrée. Avec le modèle (4.22), il y a bijection entre $I(L;X)$ et σ. Choisir la valeur de $I(L;X)$ détermine la valeur σ. Avec le modèle (4.23), choisir la valeur de $I(L;X)$ détermine la valeur du produit $\alpha\sigma$; il y a donc une infinité de couples $(\alpha;\sigma)$ solutions. Nous pouvons nous demander si tous les couples $(\alpha;\sigma)$ solutions conduisent à la même information mutuelle en sortie. Nous apportons un élément de réponse à l'aide d'un exemple.

■ **Exemple 4.8** Soit un turbo-code série (SCTC) obtenu par concaténation d'un code $(5,7)_8$ avec un code de générateur $\frac{1}{1+D}$. On considère une modulation BPSK et un canal AWGN. Nous étudions les deux décodeurs séparément. L'*a priori* est généré en utilisant le modèle (4.23) pour différentes valeurs de α tout en conservant l'information mutuelle en entrée constante. On trace sur la figure 4.2, l'information mutuelle en sortie en fonction de α. On constate que le choix de α n'est pas

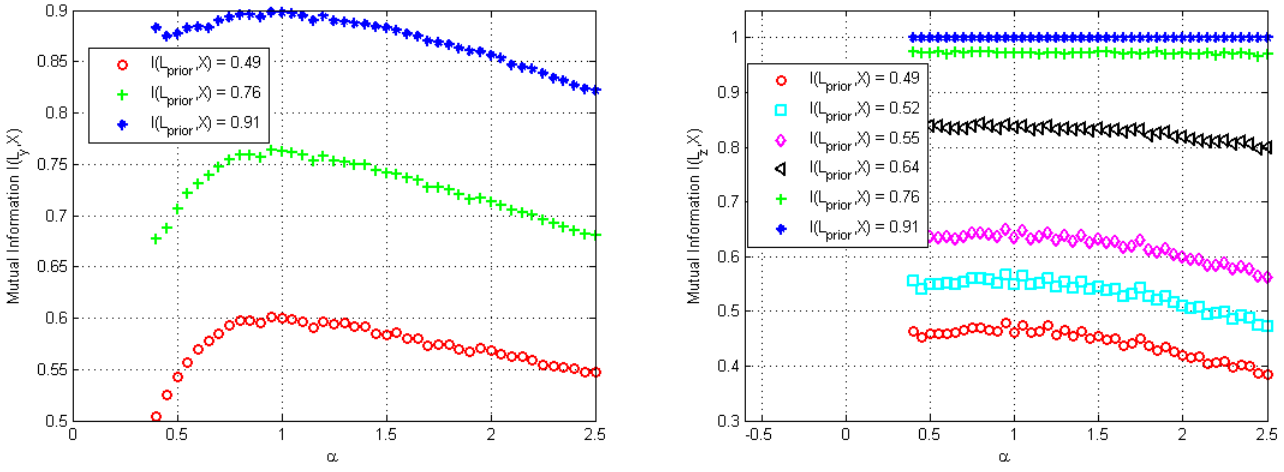


FIGURE 4.2 – SCTC $EbN0 = 2dB$ - Gauche : décodeur 1, code $\frac{1}{1+D}$ - Droite : décodeur 2, code $(5,7)_8$

neutre vis à vis de l'information mutuelle en sortie. Choisir $\alpha = 1$ est un bon choix pour les deux décodeurs testés. C'est l'optimum pour le décodeur de gauche et l'un des optimum pour le décodeur de droite [4]. ∎

A partir de cet exemple, il semble donc justifié de générer, dans les EXIT-charts, des *a priori* à partir du modèle (4.22). Les performances ainsi simulées seront conformes aux performances du système à condition que, au niveau du récepteur, on garantisse également l'appartenance des *a priori* à l'ensemble $G_1(X)$. Il semble aussi que ce soit les meilleures performances que l'on puisse obtenir. Il est donc nécessaire de déterminer au niveau du récepteur la valeur du paramètre α de chaque *a priori*.

4.6.2 Définitions et propriétés

Nous nous intéressons dans cette section à l'information mutuelle $I(L_y;L_z)$. L'objectif est de montrer le lien entre cette information mutuelle et l'information usuelle $I(L;X)$. L'objectif est aussi de disposer d'une métrique faisant intervenir des quantités connues au niveau du récepteur (ce qui n'est pas le cas de $I(L;X)$) et de caractériser le comportement du système dans son ensemble.

Définition 4.6.1 L'information mutuelle entre extrinsèques s'écrit :

$$I(L_y,L_z) = \frac{1}{\log(2)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{L_y,L_z}(\ell_y,\ell_z) \log\left(\frac{p_{L_y,L_z}(\ell_y,\ell_z)}{p_{L_y}(\ell_y) p_{L_z}(\ell_z)} \right) d\ell_y d\ell_z \tag{4.26}$$

En utilisant la propriété (4.6.1), nous pouvons montrer [119], [14] que $I(L_y,L_z)$ peut s'exprimer sous la forme :

$$I(L_y,L_z) = I(L_y,X) + I(L_z,X) - I(L_y+L_z,X) \tag{4.27}$$

On retrouve les deux informations mutuelles tracées dans les EXIT-charts que l'on sait parfaitement évaluer (hors-ligne) à l'aide de la méthode des histogrammes [106]. A l'aide de cette expression, nous pouvons en déduire les propriétés listées ci-après.

Propriété 4.6.2 Si $I(L,X)$ peut s'exprimer comme une fonction f_I d'un paramètre unique σ alors $I(L_y,L_z)$ est une fonction de deux paramètres et a pour expression :

$$I(L_y,L_z) = f_I(\sigma_y) + f_I(\sigma_z) - f_I(u(\sigma_y;\sigma_z))$$

où u est une fonction de $\mathbb{R}^+ \times \mathbb{R}^+$ vers $\mathbb{R}^+$. Si on suppose également que f_I est une fonction strictement croissante et que $u(\sigma_y;\sigma_z) \geq \max(\sigma_y;\sigma_z)$, on a

$$\max(I(L_y,X);I(L_z;X)) \leq I(L_y+L_z;X) \leq I(L_y;X) + I(L_z;X)$$

$$I(L_y,L_z) \leq \min(I(L_y,X);I(L_z;X))$$

Si $I(L_y,L_z) = m$ alors $I(L_y,X) \geq m$ et $I(L_z,X) \geq m$

Avec les modèles (4.22) et (4.23), la fonction $f_I(.)$ est la fonction $J(.)$ définie dans (4.24). C'est une fonction strictement croissante. Pour les deux modèles, $u(\sigma_y;\sigma_z) = \sqrt{\sigma_y^2 + \sigma_z^2}$. On a donc bien $u(\sigma_y;\sigma_z) \geq \max(\sigma_y;\sigma_z)$.

4. Dans un turbo-code série, seul le premier décodeur (celui de gauche ici) reçoit comme entrée les log ratio calculés à partir de la séquence reçue qui appartiennent à $G_1(X)$ pour un canal gaussien.

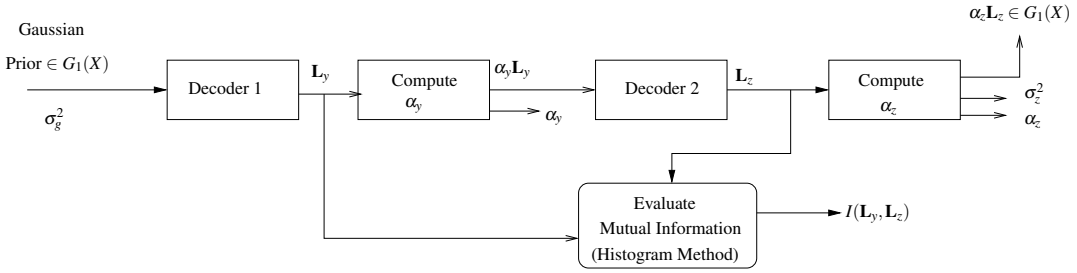


FIGURE 4.3 – SC-EXIT Chart for turbo-codes

La propriété 4.6.2 montre que l'information mutuelle entre extrinsèques est une borne inférieure de l'information mutuelle extrinsèque/message [5]. Par conséquent si, $I(L_y,L_z) = 1-\varepsilon$, on pourra affirmer que $I(L_y;X)$ et $I(L_z;X)$ appartiennent à l'intervalle $[1-\varepsilon;1]$. Réciproquement, en utilisant (4.29), si $I(L,X) \geq 1-\varepsilon$ pour $L = L_y$ et pour $L = L_z$ alors $I(L_y,L_z) \geq (1-\varepsilon)^2$. $I(L_y,L_z)$ est donc, comme $\widetilde{\mathscr{C}}$, une métrique qui peut être utilisée comme test d'arrêt ou détecteur d'erreur au niveau du récepteur à condition de pouvoir mesurer $I(L_y,L_z)$ au cours du processus itératif de décodage.

4.6.3 Estimation hors-ligne de $I(L_y,L_z)$

SC-EXIT pour turbo-codes

Si nous voulons utiliser $I(L_y,L_z)$ comme critère de performance, nous devons simuler le décodeur dans son ensemble. Nous considérons donc ici un EXIT-chart à courbe unique (SC-EXIT) [68] dont le principe est schématisé sur la figure 4.3 pour un turbo-code. Ce schéma s'étend sans difficulté à un BICM. On peut noter que d'autres variantes sont possibles puisque l'information mutuelle est une fonction d'un seul paramètre. On pourrait donc aussi tracer l'évolution de ce paramètre au lieu de tracer l'évolution de l'information mutuelle [178].

La structure est conçue afin que chaque *a priori* soit dans l'ensemble $G_1(X)$ conformément aux observations de la section 4.6.1. Dans un EXIT-chart classique, les phénomènes oscillants ou la convergence vers un minima local sont caractérisés par le croisement des courbes obtenues pour chaque décodeur. Ici, ces phénomènes se caractériseront par un passage de la courbe $\sigma_z = f(\sigma_g)$ en dessous de la 1ère bissectrice [68]. L'évaluation en parallèle de $I(L_y,L_z)$ (non considérée dans [68]) permettra de savoir pour quelle valeur de l'information mutuelle le phénomène se produit et pourra être utilisé pour construire un test d'arrêt adapté au système. Nous avons exposé la méthode dans **[14]**, elle ne sera pas détaillée ici. Nous pouvons également obtenir, à partir du SC-EXIT, une estimation du coefficient α à appliquer à chaque *a priori*. Dans le cas d'un turbo code à deux éléments, on évalue deux coefficients α_y et α_z correspondant respectivement à la pondération optimale des *a priori* L_y et L_z. Dans un Exit-Chart, X est connu ; l'évaluation de α_y et α_z est alors triviale à partir de (4.23).

5. Une borne plus fine peut être donnée sur $I(L_y,L_z)$ mais elle ne nous sera pas utile ici. Il est en effet montré dans [119] que

$$I(L_y,L_z) \leq 1-h\Big([1-h^{-1}(1-I(L_y;X))]h^{-1}(1-I(L_z;X)) + [1-h^{-1}(1-I(L_z;X))]h^{-1}(1-I(L_y;X))\Big) \quad (4.28)$$

où $h(.)$ est l'entropie et $h^{-1}(.)$ son inverse à valeur dans $[0;0.5]$. Il est également montré dans [119] que

$$I(L_y,L_z) \geq I(L_y;X)I(L_z;X) \quad (4.29)$$

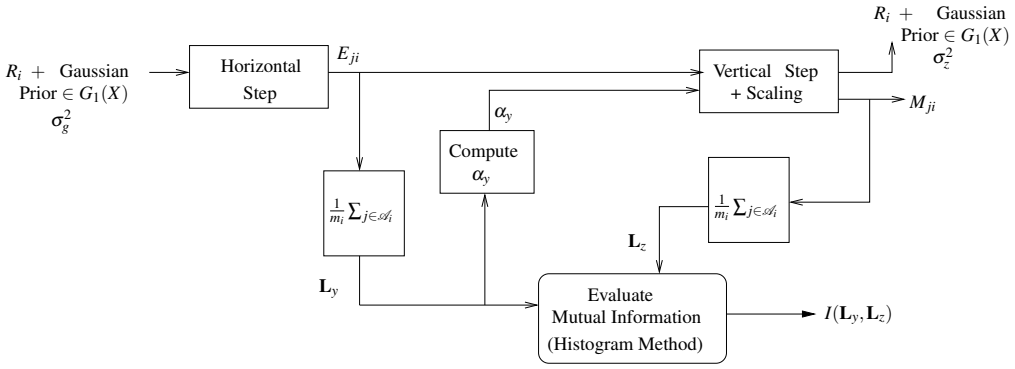


FIGURE 4.4 – SC-EXIT Chart pour LDPC

SC-EXIT pour LDPC

Pour un code LDPC, l'échange d'information a lieu entre noeuds de parité et noeuds de variable. Nous utiliserons les notations ci-dessous :

Notation 4.8. *On appelle :*
- *R_i la LLR sur le bit i calculée à partir des observations.*
- *E_{ji} l'extrinsèque (LLR) propagée du noeud de parité j vers le noeud de variable i.*
- *M_{ji} l'extrinsèque (LLR) transmise du noeud de variable i vers le noeud de parité j.*

Les quantités E_{ji} et M_{ji} sont définies uniquement si le noeud de parité j et le noeud de variable i sont connectés dans le graphe de Tanner.

Pour un algorithme *sum-product*, comme pour un algorithme *min-sum*, les messages M_{ji} sont calculés selon :

$$M_{ji} = \sum_{j' \neq j} E_{j'i} + R_i \tag{4.30}$$

Notons [6] α la pondération des extrinsèques E_{ji}. Après pondération, on a $\alpha E_{j,i} \in G_1(X)$. L'ensemble $G_1(X)$ est stable vis à vis de l'addition, comme $R_i \in G_1(X)$ on a $\alpha \sum_{j' \neq j} E_{j'i} + R_i \in G_1(X)$. Il est donc inutile de pondérer M_{ji}. Par conséquent, pour un code LDPC, seul le coefficient de pondération à appliquer à E_{ji} sera calculé. Le schéma de principe du SC-EXIT pour les LDPC est donné sur la figure 4.4.

Résultats numériques

Les valeurs obtenues pour α_y et α_z dans le cas du SCTC défini dans l'exemple 4.8 sont tracées sur la figure 4.5. L'évolution de α dans le cas du code LDPC (3,6) ainsi que dans le cas du code (4,8) est tracée dans la figure 4.6.

A partir des divers tracés, nous faisons les observations suivantes :
- La valeur de α est liée, comme nous le supposions au début de ce chapitre, à la quantité d'information acquise par le système et mesurée ici par $I(L_y, L_z)$.
- Pour tous les systèmes considérés, la valeur de α est inférieure à 1 pour les plus faibles valeurs de l'information mutuelle et passe au dessus de 1 à partir d'un certain seuil d'information. Ce seuil dépend du système et de l'algorithme choisi pour réaliser le décodage.

6. Il a été montré dans [200, 204] que le coefficient de pondération α dépend du degré du noeud de parité j. On devra donc en général calculer $\alpha_{deg(j)}$. Dans le cas d'un code régulier, tous les noeuds ont le même degré donc $\alpha = \alpha_{deg(j)}$. On se place dans ce dernier cas pour détailler la méthode proposée.

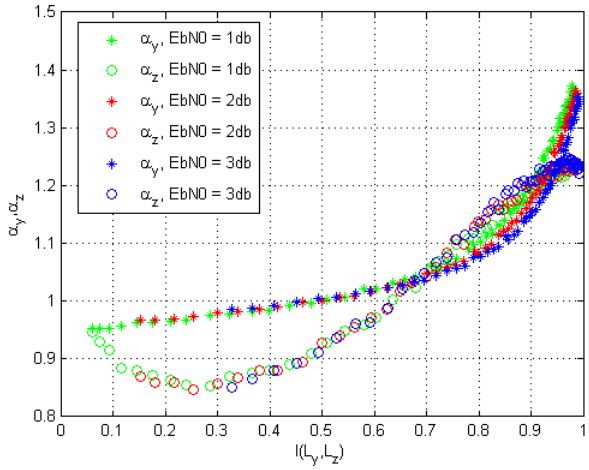


FIGURE 4.5 – α_y et α_z vs $I(L_y, L_z)$ (α_y resp. α_z pondération de l'*a priori* pour le code $(5,7)_8$ resp. $\frac{1}{1+D}$)

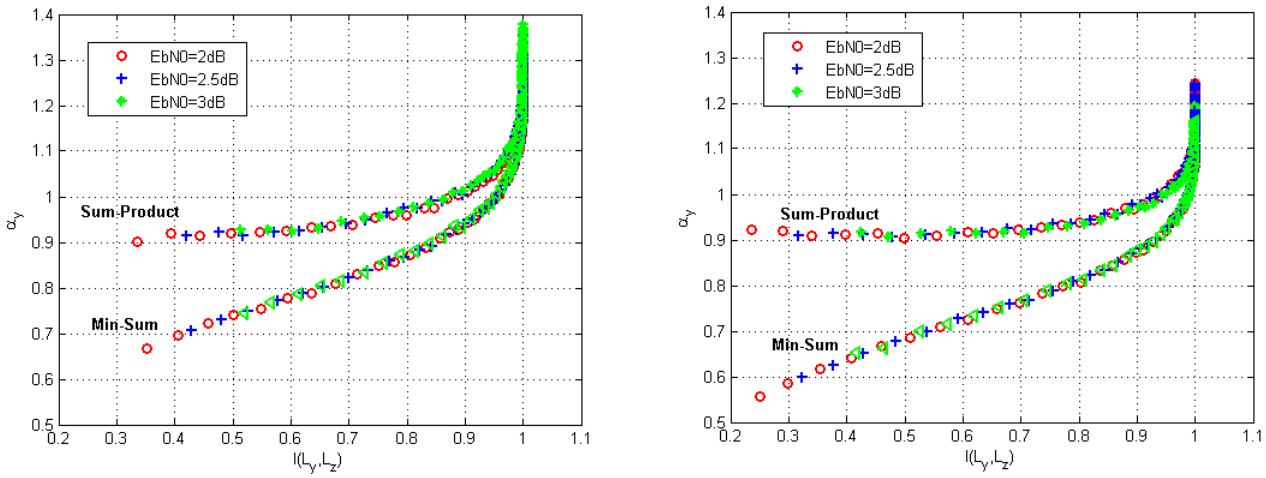


FIGURE 4.6 – SC-EXIT Chart, α vs $I(L_y, L_z)$ pour un code (3-6) LDPC (gauche) et un code (4-8) LDPC (droite)

— Sans surprise, on obtient des valeurs voisines de 1 pour le SCTC et pour l'implémentation *sum-product* alors que l'on obtient des valeurs nettement plus faibles avec l'implémentation *min-sum*. Ce phénomène est connu et de nombreux travaux ont considéré un post-traitement linéaire des LLR au niveau du récepteur. En général, le post-traitement linéaire consiste à déterminer un α optimal et indépendant des itérations [19, 48, 204]. La méthode a l'avantage d'être simple et améliore significativement les performances de l'algorithme *min-sum*. Elle est toutefois sous-optimale puisque l'on voit sur les courbes présentées ici que la valeur de α évolue au cours du processus itératif. Dans toutes les contributions citées, et dans d'autres, il est dit que α dépend du rapport signal à bruit ($EbN0$) ainsi que de l'itération. L'optimisation multi-critère est complexe à mettre en oeuvre c'est ce qui explique qu'une solution non-adaptative ait souvent été préférée. Le travail présenté ici amène une réponse à ce problème. Certes α dépend de la valeur de $EbN0$ et de l'itération mais l'information mutuelle également. Ce que montrent les courbes obtenues pour l'ensemble des système

considérés c'est que α peut s'écrire comme une fonction de la seule variable $I(L_y, Lz)$. C'est un résultat nouveau. Nous considérerons donc dans la suite le problème sous la forme

$$\alpha = F_S\Big(I(L_y, L_z)\Big) \tag{4.31}$$

où $F_s(.)$ est une fonction qui dépend du système considéré et qui peut être déterminée facilement par un EXIT-chart à courbe unique. Si nous sommes capables d'évaluer $I(L_y, L_z)$ avec fiabilité au niveau du récepteur alors nous avons trouvé un moyen simple de régler de manière adaptative la valeur de α au niveau du récepteur. Le calcul de $I(L_y, L_z)$ à partir de séquences de longueur finie est détaillé dans la section suivante.

4.6.4 Estimation en-ligne de $I(L_y; L_z)$

Nous nous intéressons ici à l'évaluation de l'information mutuelle moyennée sur l'ensemble des séquences présentes au niveau du récepteur. Nous considérons dans un premier temps Q séquences de longueur B telles que $n = BQ$. Nous définissons l'information mutuelle moyenne selon :

$$I_M := \frac{1}{Q}\sum_{k=1}^{Q} I(L_{y^k}, L_{z^k}) \tag{4.32}$$

Nous montrons le résultat suivant

Proposition 4.6.1 Si l'ensemble des v.a. intervenant dans le calcul de I_M sont ergodiques alors

$$I_M = 1 + \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \log_2\left(\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right) \tag{4.33}$$

Démonstration. Par définition,

$$I_M = \frac{1}{Q}\sum_{k=1}^{Q} E\left[\log_2\left(\frac{p(\ell_{y,k}, \ell_{z,k})}{p(\ell_{y,k})p(\ell_{z,k})}\right)\right]$$

La probabilité jointe $p(\ell_{y,k}, \ell_{z,k})$ s'écrit $p(\ell_{y,k}, \ell_{z,k}) = \sum_{x_k} \frac{p(x_k|\ell_{y,k})p(x_k|l_{z,k})}{p(x_k)} p(\ell_{y,k})p(\ell_{z,k})$. Avec $p(x_k) = \frac{1}{2}$, on en déduit

$$I_M = \frac{1}{Q}\sum_{k=1}^{Q} E\left[\sum_{x_k} \log_2\left(2\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right)\right]$$

En supposant les v.a. ergodiques, on peut écrire $I_M = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \log_2\left(2\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right)$ ■

Avec une séquence unique de longueur finie égale à n, nous pourrons estimer I_M et donc *in fine* $I(L_y; L_z)$ par

$$I(L_y; L_z) \approx 1 + \frac{1}{n}\sum_{k=1}^{n} \log_2\left(\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right) \tag{4.34}$$

Le terme $\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})$ est la constante de normalisation à appliquer pour normaliser la probabilité *a posteriori* à partir du produit des probabilités extrinsèques normalisées. Le seul cas où cette constante de normalisation vaut 1 correspond à la situation où $p(x_k|\ell_{y,k}) = p(x_k|\ell_{z,k}) = \delta x_k, \widehat{x}_{0,k}$ pour tout k. L'information mutuelle est alors maximale. On retrouve à nouveau comme résultat

que si la probabilité *a posteriori* obtenue à l'issue du processus itératif est une mesure de Dirac alors la solution obtenue donne une estimation exacte du message transmis. A partir des LLR, nous pourrons donner une nouvelle expression de l'estimateur de $I(L_y;L_z)$

$$\widehat{I}(L_y;L_z) = 1 + \frac{1}{n}\sum_{k=1}^{n} \log_2\left(\frac{1+e^{\ell_{y,k}+\ell_{z,k}}}{(1+e^{\ell_{y,k}})(1+e^{\ell_{z,k}})}\right) \tag{4.35}$$

où ℓ_z et ℓ_y appartiennent à $G_1(X)$. Nous pourrons évaluer cette quantité de manière très simple au niveau du récepteur selon

$$\widehat{I}(L_y;L_z) = 1 + \frac{1}{\log(2)n}\sum_{k=1}^{n} g(\ell_{y,k}+\ell_{z,k}) - g(\ell_{y,k}) - g(\ell_{z,k}) \tag{4.36}$$

avec $g(\ell) = \max(0,\ell) + \log(1+exp(-|\ell|))$ et où $\log(1+exp(-|\ell|))$ peut être pré-calculé pour un nombre déterminé de valeurs de ℓ.

L'estimateur $\widehat{I}(L_y;L_z)$ est calculé à partir de LLR pondérées ($\in G_1(X)$). Nous souhaitons utiliser cet estimateur, dans le processus itératif, pour déduire le facteur de pondération à appliquer à la dernière LLR calculée (ℓ_y par exemple). Par construction, $\ell_y \notin G_1(X)$. L'impact sur la valeur de $\widehat{I}(L_y;L_z)$ est évalué dans la figure 4.7. Nous considérons un LDPC (4,8) avec implémentation *min-sum* pour lequel la dynamique sur la valeur de α est la plus importante comparativement aux autres systèmes étudiés dans ce chapitre. Nous utilisons un SC-EXIT chart, et nous comparons la valeur donnée par la méthode de l'histogramme pour $I(L_y;L_z)$ (qui donne une valeur de l'information mutuelle identique quelque soit la valeur de α) à celle obtenue pour $\widehat{I}(L_y;L_z)$ calculé avec la relation (4.36) dans laquelle $\ell_z \in G_1(X)$ et $\ell_y \notin G_1(X)$. Sur la figure de gauche est tracée l'évolution de $\widehat{I}(L_y;L_z)$ en fonction de $I(L_y;L_z)$ (histogramme). Sur la figure de droite, la valeur optimale du facteur de pondération est tracée en fonction de $I(L_y;L_z)$ (en vert) et de $\widehat{I}(L_y;L_z)$ (en rouge). On constate

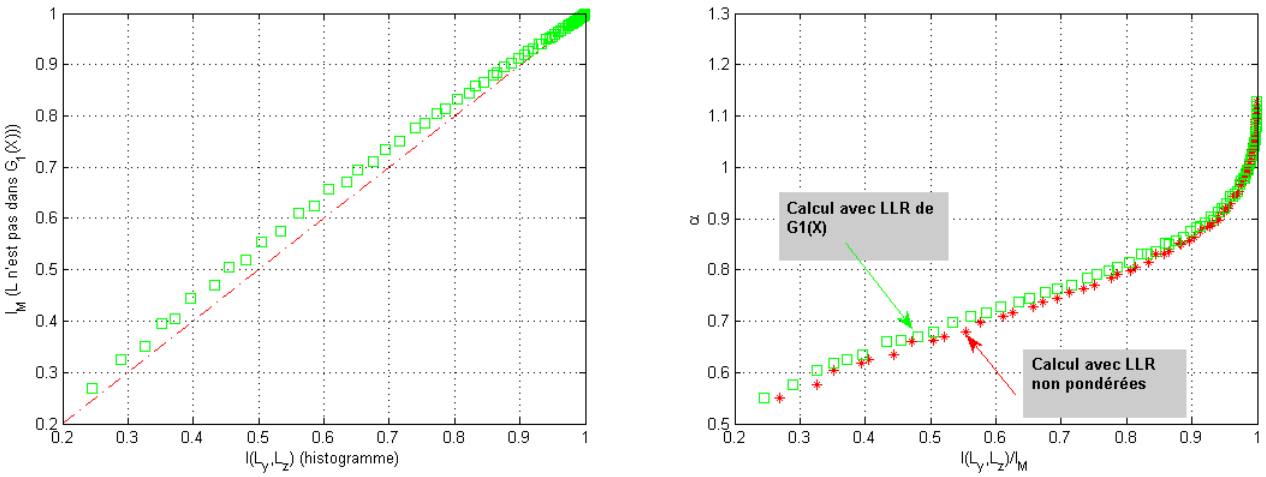


FIGURE 4.7 – (4,8) LDPC avec *min-sum*, validité du calcul de $\widehat{I}(L_y;L_z)$

que la différence entre les deux valeurs mesurées (histogramme et $\widehat{I}(L_y;L_z)$) est suffisamment faible pour que nous puissions utiliser $\widehat{I}(L_y;L_z)$ afin de calculer adaptativement α. La méthode est résumée ci-dessous.

Résultat 4.6.2 Evaluation adaptative de α_y (méthode identique pour α_z)

— *Etape 1 (Hors ligne, SC-EXIT)* :

Tracer α en fonction de $I(L_y;L_z)$ ou de $\widehat{I}(L_y;L_z)$
En déduire l'équation de $F_S(.)$ telle que $\alpha_y = F_S(\widehat{I}(L_y;L_z))$

— *Etape 2 (Au niveau du récepteur)* :

Evaluer $\widehat{I}(L_y;L_z)$ via (4.36) avec $L_z \in G_1(X)$
Calculer $\alpha_y = F_S(\widehat{I}(L_y;L_z))$
$\ell_y \leftarrow \alpha_y \ell_y$

On pourra noter que l'étape 1 doit être réalisée une seule fois pour un système donné puisqu'elle est indépendante de $EbN0$. L'étape 2 est réalisée deux fois par itération avec un turbo-code et une seule fois par itération avec un code LDPC.

4.7 Simulations

Nous terminons ce chapitre en illustrant la méthode d'estimation proposée sur un exemple. Nous appliquons notre procédure au code décrit ci-après.

■ **Exemple 4.9** On considère un code LDPC irrégulier de taux $\frac{1}{2}$ et de longueur 10032 et tel que

$$\begin{aligned}\rho(x) &= 0.25x^4 + 0.75x^{14} \\ \lambda(x) &= 0.1917x + 0.0125x^2 + 0.0417x^4 + 0.0750x^5 + 0.2041x^6 + 0.0667x^7 + 0.02250x^8 \\ &\quad + 0.0833x^9 0.0458x^{10} + 0.0541x^{12}\end{aligned}$$

■

Nous avons choisi ce code afin de pouvoir comparer la méthode proposée aux résultats récents obtenus dans [200] où une méthode de pondération adaptative des LLR est également proposée. Le code choisi présente des noeuds de parité avec deux degrés différents 5 et 15. Nous montrons dans la figure 4.8, l'évolution du facteur de pondération α_5 respectivement α_{15} à appliquer à E_{ji} où j correspond à un noeud de parité de degré 5 respectivement 15.

L'objectif est ici de trouver la meilleure pondération des LLR dans le cadre d'une implémentation *min-sum* au niveau du décodeur afin d'obtenir des performances proches de celles obtenues avec une implémentation *sum-product*.

— *Fig. 4.8, graphique de droite*. Dans [200] comme dans la plupart des publications sur le sujet, on considère que le facteur de pondération dépend de l'itération et du rapport signal à bruit. Dans [200], le deuxième paramètre (le rapport signal à bruit) est fixé à $EbN0 = (EbN0)_{threshold}$ où $(EbN0)_{threshold}$ est le seuil prédit par évolution de densité [47]. Le paramètre α devient alors une fonction d'une seule variable : le nombre d'itérations. La valeur optimale de α est ensuite déterminée par résolution du problème d'optimisation ci-dessous :

$$\alpha_j^{(it)} = \arg\min_{\alpha} \frac{1}{L} \sum_{m \in M_d(j)} \sum_{n \in N(m)} \log_2\left(1 + e^{-\alpha E_{mn}^{(it)}}\right) \tag{4.37}$$

où it est le numéro de l'itération, $M_d(j)$ est l'ensemble des noeuds de parité de degré j, $N(m)$ est l'ensemble des noeuds de variable contenus dans l'équation de parité m et où L est le nombre de termes $E_{mn}^{(it)}$ impliqués dans l'équation. Le problème (4.37) est résolu, dans [200], par recherche exhaustive dans l'intervalle $]\alpha_j^{(it-1)}; 1]$. Ce problème est équivalent à une maximisation de l'information mutuelle $I(L;X)$ où L suit le modèle (4.23) et où la moyenne statistique a été remplacée par une moyenne empirique. Les valeurs obtenues sont données dans la figure 4.8 sur le graphique de droite (graphique issu de [200]). On obtient ainsi

l'évolution de α_5 (noté α_2 sur le graphe) et de α_{15} (noté α_1 sur le graphe) en fonction de l'itération pour $EbN0 = (EbN0)_{threshold}$. La troisième courbe est la valeur de α obtenue en considérant l'ensemble des noeuds de parité. Logiquement, cette courbe est située entre les deux autres courbes.

— *Fig. 4.8, graphique de gauche*. On trace l'évolution de α_5 et de α_{15} en fonction de l'information mutuelle. Les deux courbes sont obtenues à l'aide d'un EXIT-chart comme expliqué dans la section 4.6.3 (voir le schéma de principe sur la figure 4.4).

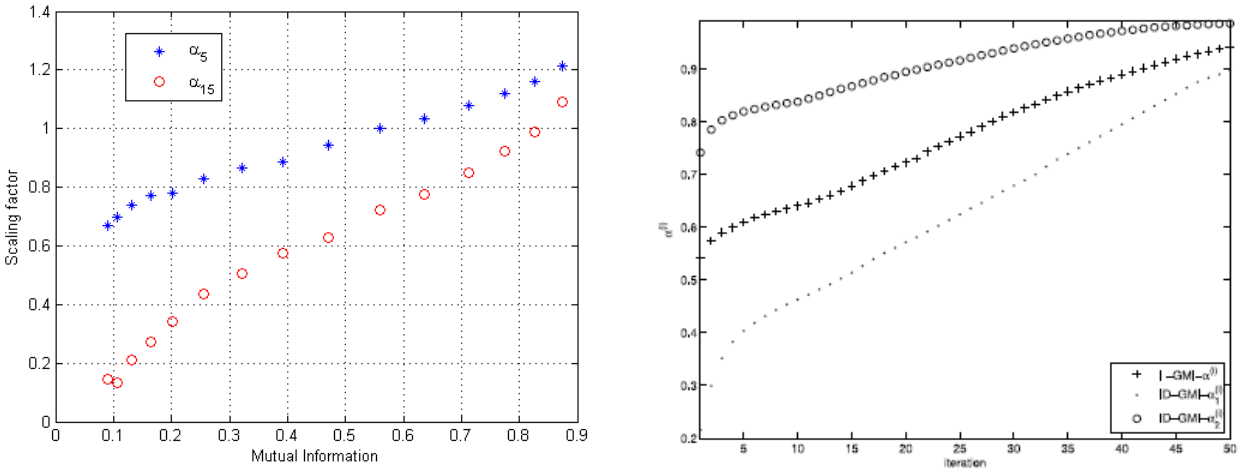


FIGURE 4.8 – Evolution de α_5 et α_{15} avec l'information mutuelle (gauche) avec l'itération (droite).

On trouve des courbes d'allure similaire sur les deux graphiques. La différence majeure est que les courbes du graphique de gauche sont valides pour n'importe quel $EbN0$ alors que celles de droites sont valides pour $EbN0 = (EbN0)_{threshold}$. Ainsi, en utilisant les courbes du graphique de droite dans un contexte à fort SNR, on risque de ralentir l'algorithme puisque les valeurs du facteur de pondération vont être sous-estimées par rapport à la confiance que l'on pourrait accorder aux LLR. En revanche, avec le graphique de gauche, le facteur de pondération sera ajusté en fonction de l'information mutuelle qui traduira convenablement la confiance à accorder aux LLR. Pour confirmer cette hypothèse, nous comparons en terme de BER et de nombre d'itérations les méthodes suivantes :

— *SP :* Algorithme *sum-product* sans correction des LLR.
— *MS avec $\alpha = f(IM)$:* Algorithme *min-sum* où α_5 et α_{15} sont déterminés à l'aide du graphique de gauche de la figure 4.8. Par ajustement polynomial, on obtient :

$$\begin{aligned} \alpha_5 &= 0.41554 + 1.9557 I_M - 2.9752 I_M^2 + 1.8319 I_M^3 \\ \alpha_{15} &= -0.23184 + 2.7474 I_M - 3.6899 I_M^2 + 2.268 I_M^3 \end{aligned}$$

où $I_M = \widehat{I}(L_y; L_z)$. On utilise ensuite l'étape 2 du résultat 4.6.2.
— *MS avec $\alpha = f(EbN0_{threshold}, it)$:* Algorithme *min-sum* où α_5 et α_{15} sont déterminés à l'aide du graphique de droite de la figure 4.8.
— *DNMS :* Algorithme *min-sum* où $\alpha_5 = 0.88$ et $\alpha_{15} = 0.68$. Ce couple constitue le meilleur couple de valeurs (non variables). Il est déterminé par simulation dans [200].
— *MS avec $\alpha = 1$:* Algorithme *min-sum* sans correction des LLR.

Les résultats sont montrés sur la figure 4.9 pour laquelle une modulation BPSK a été utilisée. Nous montrons également sur la figure 4.10 les résultats obtenus avec une modulation 256QAM. On constate dans les deux cas que la méthode que nous avons proposée a les meilleures performances en terme de BER parmi l'ensemble des implémentations de type *min-sum* considérées. Par rapport à

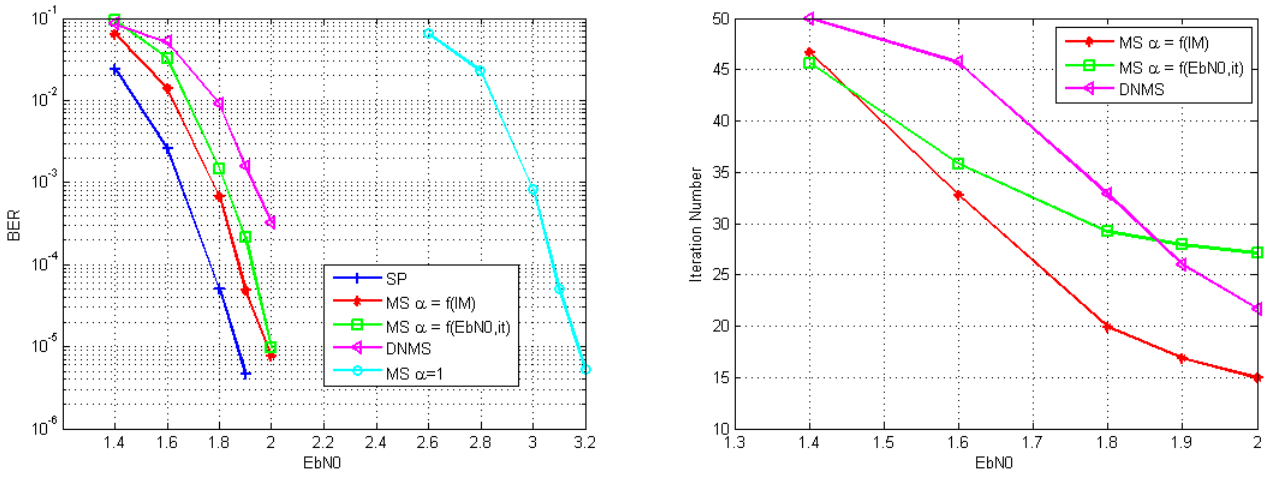


FIGURE 4.9 – Comparaison de divers algorithmes en termes de BER (gauche) et de nombre d'itérations (droite) - Modulation BPSK.

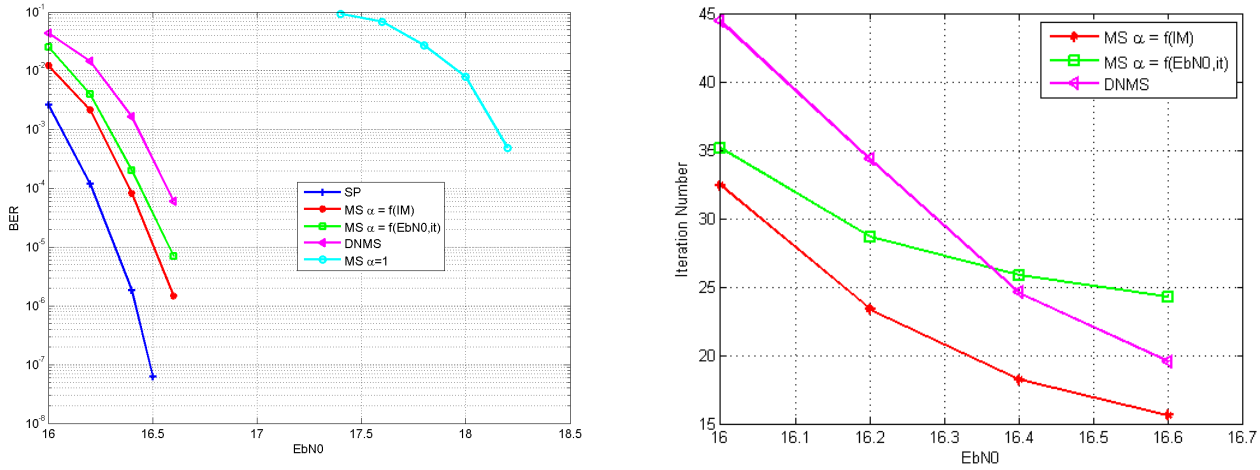


FIGURE 4.10 – Comparaison de divers algorithmes en termes de BER (gauche) et de nombre d'itérations (droite) - Modulation 256QAM.

la méthode proposée dans [200], l'amélioration est toutefois mineure. En revanche, le gain principal est obtenu sur le nombre d'itérations. La méthode proposée dans [200] donne de bons résultats pour une seule valeur de $EbN0$ (celle qui a été choisi pour faire l'acquisition des séquences α_5 et α_{15} de la figure 4.8 à droite). Lorsque $EbN0$ est supérieur à cette valeur, la convergence est ralentie car les valeurs optimales de α_5 et α_{15} sont sous-estimées. C'est cela qui est mis en évidence dans les figures 4.9 et 4.10 dans lesquelles nous observons que la méthode proposée permet d'économiser près d'un tiers des itérations dans la zone en cascade.

4.8 Conclusion

J'ai présenté dans ce chapitre un ensemble de résultats donnant une interprétation alternative du procédé de décodage itératif à l'aide de concepts issus de la théorie des jeux. Une nouvelle métrique, l'information mutuelle entre extrinsèque, a également été proposée permettant d'exacerber la capacité du récepteur itératif à faire ces propres choix en fonction de la connaissance acquise au

fil des itérations et traduite de manière chiffrée par l'information mutuelle. Les travaux décrits ici sont assez généraux et s'appliquent au BICM, turbo-codes ou encore aux codes LDPC. Des prolongements de ces travaux peuvent s'imaginer dans différentes directions. On peut, par exemple, considérer l'extension à des codes non-binaires [165]. L'information mutuelle entre extrinsèques peut en effet être calculée avec un principe similaire pour des codes binaires et non-binaires. Il a également été montré dans [28] que, sous approximation gaussienne, la distribution des LLR de dimension $q-1$ pour un code LDPC sur $GF(q)$ est décrite par un paramètre unique comme dans le cas des codes binaires. Il est donc possible de faire une étude similaire à celle présentée dans la partie 4.6. Il est en revanche difficile de prévoir si pour les codes non-binaire la pondération des LLR s'avèrerait aussi utile que pour les codes binaires. Les mêmes remarques s'appliquent à un turbo-détecteur [178]. La métrique que nous avons introduite ici peut aussi permettre de choisir un relais dans une situation de coopération ou encore pour retransmettre à l'émetteur, par un canal de retour, une mesure de l'information présente au niveau du récepteur. Ce dernier point est développé dans le chapitre 3 dans un contexte de communication sécurisée.

BILAN (THÈSE/PUBLICATION)

- 1 thèse soutenue (Z. Naja),
- 2 publication dans des revues internationales avec comité de lecture (*IEEE Trans. on Signal Processing* [R.3] et *IEEE Trans. on Communications* [R1]),
- 6 publications dans les actes de conférences internationales avec comité de lecture ([C.4], [C.6], [C.7] et [C.9] à [C.11]),
- 1 publication dans les actes d'une conférence nationale avec comité de lecture [N.2].

Bibliographie

[1] F. ABDELKEFI. “Les codes Reed-Solomon complexes pour la correction des erreurs impulsives dans les systèmes multiporteuses”. Thèse de doctorat. Paris, France : Ecole Nationale Supérieure des Télécommunications (ENST), 2002.

[2] F. ABDELKEFI, P. DUHAMEL et F. ALBERGE. “Impulse noise correction in Hiperlan 2 : improvement of the decoding algorithm and application to PAPR reduction”. In : *Proc. of ICC (International Conference on Communication)*. 2003.

[3] F. ABDELKEFI, P. DUHAMEL et F. ALBERGE. “Tests en cascade pour la correction des erreurs impulsives et la réduction du PAPR dans le contexte d’Hiperlan 2”. In : *Actes du GRETSI*. 2003.

[4] F. ABDELKEFI, P. DUHAMEL et F. ALBERGE. “Impulsive noise cancellation in multicarrier transmission”. In : *Communications, IEEE Transactions on* 53.1 (jan. 2005), pages 94–106.

[5] F. ABDELKEFI, P. DUHAMEL et F. ALBERGE. “A Necessary Condition on the Location of Pilot Tones for Maximizing the Correction Capacity in OFDM Systems”. In : *IEEE Transactions on Communications* 55.2 (fév. 2007), pages 356–366.

[6] F. ABDELKEFI et al. “On the use of cascade structure to correct impulsive noise in multicarrier systems”. In : *Communications, IEEE Transactions on* 56.11 (nov. 2008), pages 1844–1858.

[7] K. ABED-MERAIM, E. MOULINES et P. LOUBATON. “Prediction error method for second-order blind identification”. In : *IEEE Trans. Signal Processing* 45 (mar. 1997), pages 694–705.

[8] H. ABEIDA et al. “An EM Algorithm for Path Delay and Complex Gain Estimation of Slowly Varying Fading Channel for CPM Signals”. In : *Global Telecommunications Conference, 2009. GLOBECOM 2009*. Nov. 2009, pages 1–6.

[9] T. ADALI et S. HAYKIN. *Adaptive Signal Processing : Next Generation Solutions*. Wiley-IEEE Press, 2010.

[10] A. AGGARWAL et T.H. MENG. “Minimizing the Peak-to-Average Power Ratio of OFDM Signals Using Convex Optimization”. In : *IEEE Transactions on Signal Processing* 54.8 (août 2006), pages 3099–3110.

[11] R. AHLSWEDE et al. “Network information flow”. In : *IEEE Transactions on Information Theory* 46.4 (juil. 2000), pages 1204–1216.

[12] F. ALBERGE. “Accelerated Linear EM-MAP Algorithm for OFDM Channel Estimation”. In : *IEEE International Conference on Acoustics, Speech and Signal Processing*. Tome 3. Avr. 2007, pages.

[13] F. ALBERGE. “A game-theoretic interpretation of iterative decoding”. In : *EUSIPCO*. Tome 1. Barcelona, Spain, 29 Aug. - 2 Sept. 2011, pages 76–80.

[14] F. ALBERGE. “On Some Properties of the Mutual Information Between Extrinsics With Application to Iterative Decoding”. In : *IEEE Transactions on Communications* 63.5 (mai 2015), pages 1541–1553.

[15] F. ALBERGE, P. DUHAMEL et M. NIKOLOVA. “Adaptive solution for blind identification/equalization using deterministic maximum likelihood”. In : *IEEE Transactions on Signal Processing* 50.4 (avr. 2002), pages 923–936.

[16] F. ALBERGE, Z. NAJA et P. DUHAMEL. “From maximum likelihood to iterative decoding”. In : *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011*. Mai 2011, pages 3052–3055.

[17] F. ALBERGE, Z. NAJA et P. DUHAMEL. “Power Extrinsic Propagation for Turbo-Codes”. In : *IEEE Transactions on Signal Processing* 61.5 (mar. 2013), pages 1107–1111.

[18] F. ALBERGE, M. NIKOLOVA et P. DUHAMEL. “Blind identification/equalization using deterministic maximum likelihood and a partial prior on the input”. In : *IEEE Transactions on Signal Processing* 54.2 (fév. 2006), pages 724–737.

[19] A. ALVARADO et al. “Correcting Suboptimal Metrics in Iterative Decoders”. In : *IEEE International Conference on Communications, ICC ’09*. Juin 2009, pages 1–6.

[20] S.-I. AMARI et A. CICHOCKI. “Adaptive blind signal processing-neural network approaches”. In : *Proceedings of the IEEE* 86.10 (oct. 1998), pages 2026–2048.

[21] E. ARIKAN. “Channel Polarization : A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels”. In : *IEEE Transactions on Information Theory* 55.7 (juil. 2009), pages 3051–3073.

[22] M.J. ARROW. *Social Choices and Individual Values*. Yale University Press, 1963.

[23] M.-A. BADIU et al. “Message-passing algorithms for channel estimation and decoding using approximate inference”. In : *2012 IEEE International Symposium on Information Theory Proceedings (ISIT)*. Juil. 2012, pages 2376–2380.

[24] A.R.S. BAHAI et al. “A new approach for evaluating clipping distortion in multicarrier systems”. In : *IEEE Journal on Selected Areas in Communications* 20.5 (juin 2002), pages 1037–1046.

[25] H. BALTA et C. DOUILLARD. “On the influence of the extrinsic information scaling coefficient on the performance of single and double binary turbo codes”. In : *Advances in electrical and computer engineering* 2013.2 (2013), pages 77–84.

[26] O. BARRIENTOS et R. CORREA. “An algorithm for global minimization of linearly constrained quadratic functions”. In : *Journal of Global Optimization* 16 (2000), pages 77–93.

[27] H.H. BAUSCHKE et J.M. BORWEIN. “Legendre functions and the method of random Bregman projections”. In : *J. Convex. Anal.* 4.1 (1997), pages 27–67.

[28] A. BENNATAN et D. BURSHTEIN. “Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels”. In : *IEEE Transactions on Information Theory* 52.2 (fév. 2006), pages 549–583.

[29] A. BENVENISTE, M. MÉTIVIER et P. PRIOURET. *Adaptive algorithms and stochastic approximations*. Springer Verlag, 1990.

[30] P. BERGMANS. "Random coding theorem for broadcast channels with degraded components". In : *Information Theory, IEEE Transactions on* 19.2 (mar. 1973), pages 197–207.

[31] P. BERGMANS. "A simple converse for broadcast channels with additive white Gaussian noise (Corresp.)" In : *IEEE Transactions on Information Theory* 20.2 (mar. 1974), pages 279–280.

[32] C. BERROU, A. GLAVIEUX et P. THITIMAJSHIMA. "Near Shannon limit error-correcting coding and decoding : Turbo-codes". In : *IEEE International Conference on Communications. ICC '93 Geneva.* Tome 2. Mai 1993, 1064–1070 vol.2.

[33] A. BEUTEL et al. "Elastic distributed bayesian collaborative filtering". In : *In NIPS workshop on Distributed Machine Learning and Matrix Computations*. 2014.

[34] K.L. BLACKARD, T.S. RAPPAPORT et C.W. BOSTIAN. "Measurements and Models of Radio Frequency Impulsive Noise for Indoor Wireless Communications". In : *IEEE J. Select Areas Commun.* 11 (sept. 1993), pages 991–1001.

[35] R. E. BLAHUT. *Theory and practice of error control codes*. Reading MA : Addison-Wesley, 1987.

[36] R. E. BLAHUT. *Algebraic methods for Signal Processing and Commnications Coding*. New York : Springer-Verlag, 1992.

[37] R.E. BLAHUT. "Transform Techniques for Error Control Codes". In : *IBM Journal of Research and Development* 23 (mai 1979), pages 299–315.

[38] R.E. BLAHUT. "Algebraic fields, signal processing and error control". In : *Proc. IEEE* 73.5 (mai 1985), pages 874–893.

[39] M. BLOCH et J. BARROS. *Physical-Layer security : from information theory to security engineering*. Cambridge Univ. Press, Cambridge, U.K., 2011.

[40] I. BOMZE et G. DANNINGER. "A finite algorithm for solving general quadratic problem". In : *Journal of Global Optimization* 4 (1994), pages 1–16.

[41] D. BOSS, K-D KAMMEYER et T. PETERMANN. "Is blind channel estimation feasible in mobile communications systems ? A study based on GSM". In : *IEEE Journal on Selected Areas in Communications* 16 (oct. 1998), pages 1479–1492.

[42] S. ten BRINK. "Convergence behavior of iteratively decoded parallel concatenated codes". In : *IEEE Transactions on Communications* 49.10 (oct. 2001), pages 1727–1737.

[43] E.J. CANDES et P.A. RANDALL. "Highly Robust Error Correction byConvex Programming". In : *IEEE Transactions on Information Theory* 54.7 (juil. 2008), pages 2829–2840.

[44] E.J. CANDES, J. ROMBERG et T. TAO. "Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information". In : *IEEE Transactions on Information Theory* 52.2 (fév. 2006), pages 489–509.

[45] V. CEVHER, S. BECKER et M. SCHMIDT. "Convex Optimization for Big Data : Scalable, randomized, and parallel algorithms for big data analytics". In : *IEEE Signal Processing Magazine* 31.5 (sept. 2014), pages 32–43.

[46] M. CHAMI, M. PISCHELLA et D. LE RUYET. "Adaptive Decoding Strategy with Superposition Coding for Cognitive Radio Systems". In : *European Wireless 2014 ; 20th European Wireless Conference ; Proceedings of*. Mai 2014, pages 1–6.

[47] J. CHEN et M.P.C. FOSSORIER. “Density evolution for two improved BP-Based decoding algorithms of LDPC codes”. In : *IEEE Communications Letters* 6.5 (mai 2002), pages 208–210.

[48] J. CHEN et M.P.C. FOSSORIER. “Near optimum universal belief propagation based decoding of low-density parity check codes”. In : *IEEE Transactions on Communications* 50.3 (mar. 2002), pages 406–414.

[49] J. CHEN et al. “Reduced-Complexity Decoding of LDPC Codes”. In : *IEEE Transactions on Communications* 53.8 (août 2005), pages 1288–1299.

[50] S. CHRÉTIEN et A. O. HERO. “Kullback Proximal Algorithms for Maximum Likelihood Estimation”. In : *IEEE Trans. on Inf. Theory* 46.5 (2000), pages 1800–1810.

[51] P. G. CIARLET. *Introduction à l’analyse matricielle et à l’optimisation*. Masson, 1985.

[52] ETSI Normalization COMMITEE. *Channel Models for HiperLan type 2 in different indoor scenarios*. ETSI standard, European Telecommunications Standards Institute - 3ERI085B. Sophia-Antipolis, Valbonne, France, 1998.

[53] ETSI Normalization COMMITEE. *Broadband Radio Access Networks (BRAN) ; HYPERLAN type 2 ; System overview*. ETSI standard, ETSI TR 101 683. 2002.

[54] P. COMON et G.H. GOLUB. “Tracking a few extreme singular values and vectors in signal processing”. In : *Proceedings of the IEEE* 78.8 (août 1990), pages 1327–1343.

[55] T. COVER. “Broadcast channels”. In : *IEEE Transactions on Information Theory* 18.1 (jan. 1972), pages 2–14.

[56] T. COVER et A.E. GAMAL. “Capacity theorems for the relay channel”. In : *IEEE Transactions on Information Theory* 25.5 (sept. 1979), pages 572–584.

[57] T.M. COVER. “Comments on broadcast channels”. In : *IEEE Transactions on Information Theory* 44.6 (oct. 1998), pages 2524–2530.

[58] I. CSISZAR et F. MATÚŠ. “Information projections revisited”. In : *Information Theory, IEEE Transactions on* 49.6 (juin 2003), pages 1474–1490.

[59] E. DE CARVALHO, S. Mohamad OMAR et D.T.M SLOCK. “Performance and complexity analysis of blind FIR channel identification algorithms based on deterministic maximum likelihood in SIMO systems”. In : *Circuits, Systems, and Signal Processing* 32 (avr. 2013), pages 683–709.

[60] J. DEHAENE. “Continuous-time matrix algorithms, systolic algorithms and adaptive neural network”. Thèse de doctorat. Katholoieke Univ. leuven, Belgium, oct. 1995.

[61] J.-P. DELMAS et F. ALBERGE. “Lois asymptotiques d’estimateurs adaptatifs de sous-espaces introduits dans la littérature neuronale”. In : *Actes 16ème Colloque GRETSI*. Sept. 1997, pages 1097–1100.

[62] J.-P. DELMAS et F. ALBERGE. “Asymptotic performance analysis of subspace adaptive algorithms introduced in the neural network literature”. In : *IEEE Transactions on Signal Processing* 46.1 (jan. 1998), pages 170–182.

[63] J.P. DELMAS et J.F. CARDOSO. “Performance analysis of an adaptive algorithm for tracking dominant subspaces”. In : *IEEE Trans. on Signal Processing* (nov. 1998), pages 3045–3057.

[64] A. P. DEMPSTER, N. M. LAIRD et D. B. RUBIN. “Maximum likelihood from incomplete data via EM algorithm”. In : *J. Royal Statist. Society* 39 (1977).

[65] Z. DING. “An Outer-Product Decomposition Algorithm For Multichannel Blind Identification”. In : *Proc. 8th IEEE Workshop on Stat. Signal and Array Processing*. Juin 1996, pages 132–135.

[66] Z. DING. “Matrix outer-product decomposition method for blind multiplechannel identification”. In : *IEEE Trans. on Signal Processing* 45.12 (déc. 1997), pages 3053–3061.

[67] Z. DING et G. LI. “Single-channel blind equalization for GSM cellular systems”. In : *IEEE Journal in Selected Areas in Communications* 16 (oct. 1998), pages 1493–1505.

[68] D. DIVSALAR, S. DOLINAR et F. POLLARA. “Iterative turbo decoder analysis based on density evolution”. In : *IEEE Journal on Selected Areas in Communications* 19.5 (mai 2001), pages 891–907.

[69] C. DOUILLARD et M. JEZEQUEL. “Chapter 1 : Turbo Codes : From First Principles to Recent Standards ”. In : Library in Mobile and Wireless Communications. David Declercq (Editor), Marc Fossorier (Editor), Ezio Biglieri (Editor). Academic Press, 2014. Chapitre Channel Coding : Theory, Algorithms, and Applications, pages 1–52. ISBN : 978-0123964991.

[70] A. DOWLER, A. NIX et J. MCGEEHAN. “Data-derived iterative channel estimation with channel tracking for a mobile fourth generation wide area OFDM system”. In : *Global Telecommunications Conference*. Tome 2. Déc. 2003, 804–808 Vol.2.

[71] R.L. DYKSTRA. “An Iterative Procedure for Obtaining I-Projections onto the Intersection of Convex Sets”. In : *Ann. Probab.* 13.3 (1985), pages 975–984.

[72] P. EGGERMONT. “Multiplicative iterative algorithms for convex programming”. In : *Linear Algebra Appl.* 130 (1990), pages 25–42.

[73] H. EL GAMAL et Jr. HAMMONS A.R. “Analyzing the turbo decoder using the Gaussian approximation”. In : *IEEE International Symposium on Information Theory*. Fév. 2000, pages 319–.

[74] M. ESMAEILI, M. DAKHILALIAN et T.A. GULLIVER. “New secure channel coding scheme based on randomly punctured quasi-cyclic-low density parity check codes”. In : *IET Communications* 8.14 (sept. 2014), pages 2556–2562.

[75] M.P.C. FOSSORIER, M. MIHALJEVIĆ et H. IMAI. “Reduced complexity iterative decoding of low-density parity check codes based on belief propagation”. In : *IEEE Transactions on Communications* 47.5 (mai 1999), pages 673–680.

[76] M. FU. “Stochastic analysis of turbo decoding”. In : *IEEE Transactions on Information Theory* 51.1 (jan. 2005), pages 81–100.

[77] M. FU. “On Gaussian Approximation for Density Evolution of Low-Density Parity-Check Codes”. In : *IEEE International Conference on Communications, 2006. ICC ’06.* Tome 3. Juin 2006, pages 1107–1112. DOI : 10.1109/ICC.2006.254895.

[78] R. GALLEGO et al. “Semi-blind equalization for GMSK-based mobile communications”. In : *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Tome 4. Mai 2004, pages.

[79] R.K. GANTI et al. “Implementation and Experimental Results of Superposition Coding on Software Radio”. In : *Communications (ICC), 2010 IEEE International Conference on*. Mai 2010, pages 1–5.

[80] D. GESBERT. “Egalisation et identification multi-voies : Methodes auto-adaptatives au second ordre”. Thèse de Doctorat. ENST, Paris, 1997.

[81] D. GESBERT et P. DUHAMEL. “Robust Blind Channel Identification and Equalization based on Multi-Step Predictors”. In : *ICASSP Proc*. Avr. 1997, pages 3621–3624.

[82] D. GESBERT, P. DUHAMEL et S. MAYRARGUE. “Blind least-square approaches for joint data/channel estimation”. In : *IEEE DSP Workshop*. Sept. 1996.

[83] M. GHOSH et C.L. WEBER. “Maximum likelihood blind equalization”. In : *Opt. Eng.* 31.6 (juin 1992), pages 1224–1228.

[84] S. GOEL et R. NEGI. “Guaranteeing Secrecy using Artificial Noise”. In : *IEEE Transactions on Wireless Communications* 7.6 (juin 2008), pages 2180–2189.

[85] G. C. GOODWIN et K. S. SIN. *Adaptive Filtering, Prediction and Control*. Englewood Cliffs, NJ : Prentice Hall, 1984.

[86] D. GORENSTEIN et N. ZIERLER. “Encoding and error-correction procedures for the Bose-Chaudhuri codes”. In : *J. Soc. Ind. Appl. Math.* 9 (juin 1961), pages 207–214.

[87] P.J. GREEN. “On the use of the EM algorithm for penalized likelihood estimation”. In : *J.R. Statist. Soc. B* (1990), pages 443–452.

[88] M.I. GURELLI et C.L. NIKIAS. “A new eigenvector-based algorithm for multichannel blind deconvolution of input colored signals”. In : *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Tome 4. Avr. 1993, 448–451 vol.4.

[89] W.K. HARRISON et al. “Coding for Cryptographic Security Enhancement Using Stopping Sets”. In : *IEEE Transactions on Information Forensics and Security* 6.3 (sept. 2011), pages 575–584.

[90] C.R. HARTMANN et K.K. TZENG. “Generalizations of the BCH Bound”. In : *Information and Control* 20 (1972), pages 489–498.

[91] C.R.P. HARTMANN, K.K. TZENG et Robert T. CHIEN. “Some results on the minimum distance structure of cyclic codes”. In : *IEEE Transactions on Information Theory* 18.3 (mai 1972), pages 402–409.

[92] S. HAYKIN et B. WIDROW. *Least-Mean-Square Adaptive Filters*. Wiley-Interscience, 2003.

[93] X. HE et A. YENER. “The Role of Feedback in Two-Way Secure Communications”. In : *IEEE Transactions on Information Theory* 59.12 (déc. 2013), pages 8115–8130.

[94] G. HEINE et H. SAGKOB. *GPRS : Gateway to Third-generation Mobile Networks*. Artech House Publishers, fév. 2003.

[95] W. HENKEL. “Analog codes for peak-to-average ratio reduction”. In : *3rd ITG Conf. Source and Channel Coding*. Munich, Germany, jan. 2000.

[96] A.O. HERO et J.A. FESSLER. “Convergence in norm for alternating Expectation-Maximization (EM) type algoritms”. In : *Statistica Sinica* 5 (1995), pages 41–54.

[97] R.A. HORN et C.R. JONHSON. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[98] K. HORNIK et C.M. KUAN. “Convergence analysis of local feature extraction algorithms”. In : *Neural Networks* 5 (1992), pages 229–240.

[99] R. HORST et H. TUY. *Global Optimization (Deterministic approaches)*. Springer-Verlag, 1993.

[100] Y. HUA. “Fast maximum likelihood for blind identification of multiple FIR channels”. In : *IEEE Transactions on Signal Processing* 44.3 (mar. 1996), pages 661–672.

[101] A.S. IBRAHIM et al. “Cooperative communications with relay-selection : when to cooperate and whom to cooperate with ?” In : *IEEE Transactions on Wireless Communications* 7.7 (juil. 2008), pages 2814–2827.

[102] S. IKEDA, T. TANAKA et S.-I. AMARI. “Information geometry of turbo and low-density parity-check codes”. In : *IEEE Transactions on Information Theory* 50.6 (juin 2004), pages 1097–1114.

[103] R. A. ILTIS. “Joint Estimation of PN Code Delay and Multipath Using the Extended Kalman Filter”. In : *IEEE Trans. on Comm.* 38.10 (oct. 1990), pages 1677–1685.

[104] E. JAFFROT et M. SIALA. “Turbo channel estimation for OFDM systems on higly time and frequency selective channels”. In : *Proc. ICASSP*. Juin 2000.

[105] N. JAOUA et al. “Joint estimation of state and noise parameters in a linear dynamic system with impulsive measurement noise : Application to OFDM systems”. In : *Digital Signal Processing* 35 (2014), pages 21–36.

[106] S.J. JOHNSON. *Iterative Error-Correction*. Cambridge University Press, 2010.

[107] S. KAIMALETTU et al. “Constellation Shaping using LDPC Codes”. In : *IEEE International Symposium on Information Theory*. Juin 2007, pages 2366–2370.

[108] I. KARASALO. “Estimating the covariance matrix by signal subspace averaging”. In : *IEEE Transactions on Acoustics, Speech and Signal Processing* 34.1 (fév. 1986), pages 8–12.

[109] A. KATSIOTIS, N. KOLOKOTRONIS et N. KALOUPTSIDIS. “Secure encoder designs based on turbo codes”. In : *IEEE International conference on communications (ICC)*. London, UK, juin 2015.

[110] S.M. KAY et A.K. SHAW. “Frequency estimation by principal component AR spectral estimation method without eigendecomposition”. In : *IEEE Transactions on Acoustics, Speech and Signal Processing* 36.1 (jan. 1988), pages 95–101.

[111] M. EL-KHAMY et al. “Online log-likelihood ratio scaling for robust turbo decoding”. In : *IET Communications, IE* 8.2 (jan. 2014), pages 217–226.

[112] A. KHISTI et G. W. WORNELL. “Secure Transmission With Multiple Antennas I : The MISOME Wiretap Channel”. In : *IEEE Transactions on Information Theory* 56.7 (juil. 2010), pages 3088–3104.

[113] J. KHUN-JUSH et al. “A new approach for evaluating clipping distortion in multicarrier systems”. In : *Ericsson review* 2 (fév. 2000).

[114] D. KLINC et al. “LDPC Codes for the Gaussian Wiretap Channel”. In : *IEEE Transactions on Information Forensics and Security* 6.3 (sept. 2011), pages 532–540.

[115] C. KOMNINAKIS et al. “Multi-input multi-output fading channel tracking and equalization using Kalman estimation”. In : *IEEE Trans. on Signal Processing* 50.5 (mai 2002), pages 1065–1076.

[116] G. KRAMER, M. GASTPAR et P. GUPTA. “Cooperative Strategies and Capacity Theorems for Relay Networks”. In : *IEEE Transactions onInformation Theory* 51.9 (sept. 2005), pages 3037–3063.

[117] M. KRISTENSSON, B. OTTERSTEN et D. SLOCK. “Blind Subspace identification of a BPSK Communication Channel”. In : *Proc. of 30th Asilomar Conference on Signals, Systems and Computers*. Tome 2. Nov. 1996, pages 828–832.

[118] R. KUMARESAN. “Rank reduction technique and burst-error correction decoding in real/complex fields”. In : *Proc. of Asilomar Conference on Signals, Systems and Computers*. Nov. 1985, pages 457–461.

[119] I. LAND et al. “Bounds on Information combining”. In : *IEEE Transactions on Information Theory* 51.2 (2005), pages 612–619.

[120] S. LASAULCE, M. DEBBAH et E. ALTMAN. *Methodologies for analyzing equilibria in wireless games*. IEEE Signal Proc. Magazine, special issue on Game Theory for Signal Processing. Sept. 2009.

[121] S. LASAULCE et H. TEMBINE. *Game Theory and Learning for Wireless Networks - Fundamentals and Applications*. Académic Press, Elsevier, 2011.

[122] P. A. LAURENT. “Exact and Approximative Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses (AMP)”. In : *IEEE Trans. on Communications* 34 (1986), pages 150–162.

[123] H. LIU, Xu G. et Tong L. “A deterministic approach to blind identification of multi-channel FIR systems”. In : *1994 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Tome iv. Avr. 1994, IV/581–IV/584 vol.4.

[124] J. LIU et J. LI. “Parameter estimation and error reduction for OFDM-based WLANs”. In : *IEEE Transactions on Mobile Computing* 3.2 (avr. 2004), pages 152–163.

[125] I.M. MAHAFENO, Y. LOUET et J.-F. HELARD. “Peak-to-average power ratio reduction using second order cone programming based tone reservation for terrestrial digital video broadcasting systems”. In : *IET Communications* 3.7 (juil. 2009), pages 1250–1261.

[126] T.G. MARSHALL. “Decoding of Real-Number Sequence Error-Correction Codes”. In : *Proc. of Globecom*. Nov. 1983, pages 1299–1303.

[127] T.G. MARSHALL. “Coding of Real-Number Sequence for Error Correction : A Digital Signal Processing Problen”. In : *IEEE J. on Select. Areas in Communications* SAC-2.2 (mar. 1984), pages.

[128] B. MARTINET. “Régularisation d’inéquations variationnelles par approximations successives”. In : *Revue Francaise d’information et de Recherche Operationnelle*. Tome 3. Jan. 1970, pages 154–179.

[129] F. MARVASTI et al. “Efficient algorithms for burst error recovery using FFT and other transform kernels”. In : *IEEE Trans on Signal Proc.* 47 (avr. 1999), pages.

[130] L. MAZET et al. “An EM based semi-blind channel estimation algorithm designed for OFDM systems”. In : *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*. Tome 2. Nov. 2002, 1642–1646 vol.2.

[131] L. MAZET et al. “EM-based semi-blind estimation of time-varying channels”. In : *4th IEEE Workshop on Signal Processing Advances in Wireless Communications. SPAWC 2003*. Juin 2003, pages 205–209.

[132] R.J. MCELIECE, D.J.C. MACKAY et Jung-Fu CHENG. “Turbo decoding as an instance of Pearl’s belief propagation algorithm”. In : *IEEE Journal on Selected Areas in Communications* 16.2 (fév. 1998), pages 140–152.

[133] G. J. MCLACHLAN et T. KRISHNAN. *The EM Algorithm and Extensions*. Wiley Series in Probability et Statistics, 1997.

[134] X.L. MENG et D.B. RUBIN. “Using EM to obtain asymptotic variance-covariance matrices : the SEM algorithm”. In : *J. Amer. Statist. Assoc.* 86 (1991), pages 899–909.

[135] E.V. MEULEN. “Three-terminal communication channels”. In : *Advances in Applied Probability* 3.1 (1971), pages 120–154.

[136] Z. MHEICH, F. ALBERGE et P. DUHAMEL. “Achievable rates optimization for broadcast channels using finite size constellations under transmission constraints”. In : *EURASIP Journal on Wireless Communications and Networking* (jan. 2013), pages 1–15.

[137] Z. MHEICH, F. ALBERGE et P. DUHAMEL. “Achievable Secrecy Rates for the Broadcast Channel With Confidential Message and Finite Constellation Inputs”. In : *IEEE Transactions on Communications* 63.1 (jan. 2015), pages 195–205.

[138] Z. MHEICH et al. “Rate-adaptive secure HARQ protocol for block-fading channels”. In : *2014 Proceedings of the 22nd European Signal Processing Conference*. Sept. 2014, pages 830–834.

[139] D. MIDDLETON. “Canonical and Quasi-canonical Probability Models of Class A Interference”. In : *IEEE Trans. on Electromagn. Compat.* 26 (fév. 1984), pages 19–28.

[140] D. MIDDLETON. “Non-Gaussian Noise Models in Signal Processing for Telecommunications : New methods and results for class A and class B models”. In : *IEEE Trans. on Inform Theory* 45.4 (mai 1999).

[141] T. P. MINKA. “Expectation Propagation for Approximate Bayesian Inference”. In : *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*. UAI ’01. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2001, pages 362–369.

[142] G.J. MINTY. “Monotone (nonlinear) operators in Hilbert space”. In : *Duke Math. Journal* 29 (1962), pages 341–346.

[143] M. MOHER et T.A. GULLIVER. “Cross-entropy and iterative decoding”. In : *IEEE Transactions on Information Theory* 44.7 (nov. 1998), pages 3097–3104.

[144] A. MONTANARI et N. SOURLAS. “The statistical mechanics of turbo-codes”. In : *Eur. Phys. J. B.* 18 (2000), pages 107–119.

[145] T.K. MOON. “The Expectation-Maximization Algorithm”. In : *IEEE Signal Prossessing Magazine* (nov. 1996), pages 47–60.

[146] J.J. MORÉ. “Nonlinear generalizations of matrix diagonal dominance with application to Gauss-Seidel iterations”. In : *SIAM J. Numer. Anal.* 9.2 (1972), pages 357–378.

[147] J.J. MOREAU. “Proximité et dualité dans un espace Hilbertien”. In : *Bull.Soc.Math.France* 93 (1965), pages 273–299.

[148] E. MOULINES et al. “Subspace methods for the blind identification of multichannel FIR filters”. In : *IEEE Transactions on Signal Processing* 43.2 (fév. 1995), pages 516–525.

[149] B. MUQUET, P. DUHAMEL et A. de COURVILLE. “Geometrical interpretation of iterative turbo decoding”. In : *IEEE International Symposium on Information Theory*. Juin 2002, pages 142–145.

[150] K.G. MURTY et S.N. KABADI. “Some NP-complete Problems in quadratic and nonlinear programming”. In : *Mathematical programming* 39 (1987), pages 117–129.

[151] T.Y. AL-NAFFOURI. “An EM-Based Forward-Backward Kalman Filter for the Estimation of Time-Variant Channels in OFDM”. In : *IEEE Transactions on Signal Processing* 55.7 (juil. 2007), pages 3924–3930.

[152] T.Y. AL-NAFFOURI, A.A. QUADEER et G. CAIRE. “Impulse Noise Estimation and Removal for OFDM Systems”. In : *IEEE Transactions on Communications* 62.3 (mar. 2014), pages 976–989.

[153] Z. NAJA. “Interprétation et amélioration d’une procédure de démodulation itérative.” Thèse de doctorat. Univ. Paris-Sud, France, avr. 2011.

[154] Z. NAJA, F. ALBERGE et P. DUHAMEL. “Geometrical interpretation and improvements of the Blahut-Arimoto’s algorithm”. In : *IEEE International Conference on Acoustics, Speech and Signal Processing*. Avr. 2009, pages 2505–2508.

[155] Z. NAJA, F. ALBERGE et P. DUHAMEL. “Méthode du point proximal : principe et applications aux algorithmes itératifs”. In : *Actes du GRETSI (GRETSI’09)*. Dijon, France, sept. 2009.

[156] B. NAZER et M. GASTPAR. “Compute-and-forward : Harnessing interference with structured codes”. In : *IEEE International Symposium on Information Theory*. Juil. 2008, pages 772–776.

[157] J.M.M. OCLOO et F. ALBERGE. “OFDM Channel Estimation by a Linear EM-Map Algorithm”. In : *IEEE International Conference on Acoustics, Speech and Signal Processing*. Tome 4. Mai 2006, pages IV–IV.

[158] E. OJA. “Principal components, Minor components, and Linear Neural Networks”. In : *Neural Networks* 5 (1992), pages 927–935.

[159] M.J. OSBORNE et A. RUBINSTEIN. *A course in game theory*. MIT Press, 1994.

[160] N.L. OWSLEY. “Adaptive data orthogonalization”. In : *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP ’78.* Tome 3. Avr. 1978, pages 109–112.

[161] P. PAKZAD et V. ANANTHARAM. “Kikuchi approximation method for joint decoding of LDPC codes and partial-response channels”. In : *IEEE Transactions on Communications* 54.7 (juil. 2006), pages 1149–1153.

[162] P.M. PARDALOS et G. SCHNITGER. “Checking local optimality in constrained quadratic programming is NP-hard”. In : *Operations Research Letters* 7.1 (1988), pages 33–35.

[163] W.W. PETERSON. “Encoding and error-correction procedures for the Bose-Chaudhuri codes”. In : *Information Theory, IRE Transactions on* 6.4 (sept. 1960), pages 459–470.

[164] E. PITE et P. DUHAMEL. “Bilinear methods for blind channel equalization : (no) l06 76 78 29 17ocal minimum issues”. In : *ICASSP Proc.* Mai 1998.

[165] C. POULLIAT, M. FOSSORIER et D. DECLERCQ. “Design of regular (2,d/sub c/)-LDPC codes over GF(q) using their binary images”. In : *Communications, IEEE Transactions on* 56.10 (oct. 2008), pages 1626–1635.

[166] 3rd Generation Parternship PROJECT. *Technical Specification Group GSM/Edge, Radio Access Network, Radio transmission and reception, Release 1999 (3GPP TS 05.05 Version 8.16.0)*. Août 2003.

[167] A.A. QUADEER, M.S. SOHAIL et T.Y. AL-NAFFOURI. “A compressed sensing based method with support refinement for impulse noise cancelation in DSL”. In : *2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. Juin 2013, pages 255–259.

[168] C.R. RAO. *Linear Statistical Interference and Its Applications*. New-York, Wiley, 1973.

[169] E. Del RE et al. “Design of a digital MLSE receiver for mobile radio communications”. In : *Proc. of Globecom*. Tome 2. Déc. 1991, pages 1469–1473.

[170] G. R. REDINBO. “Decoding Real-Block Codes : Activity detection, Wiener estimation”. In : *IEEE Transactions on Information Theory* 46 (mar. 2000), pages 609–623.

[171] T. RICHARDSON. “The geometry of turbo-decoding dynamics”. In : *IEEE Trans. on Inform. Theory* 46.1 (2000), pages 9–23.

[172] E. RIEGLER et al. “Merging Belief Propagation and the Mean Field Approximation : A Free Energy Approach”. In : *IEEE Transactions on Information Theory* 59.1 (jan. 2013), pages 588–602.

[173] R.T. ROCKAFELLAR. “Monotone operators and the proximal point algorithm”. In : *SIAM Journal on Control and Optimization* 14 (1976), pages 877–898.

[174] C. ROOS. “A Generalization of the BCH bound for Cyclic Codes including the Hartmann-Tzeng Bound”. In : *Journal of Combinatorial Theory series A* 33.2 (1982), pages 229–232.

[175] H. RUTISHAUSER. “Computational aspects of F.L.Bauer’s simultaneous iteration method”. In : *Numer. Math.* 13 (1969), pages 4–13.

[176] O. SAHIN et E. ERKIP. “Achievable Rates for the Gaussian Interference Relay Channel”. In : *IEEE Global Telecommunications Conference*. Nov. 2007, pages 1627–1631.

[177] S.L. SCOTT, A.W. BLOCKER et F.V. BONASSI. “Bayes and Big Data : The Consensus Monte Carlo Algorithm”. In : *Bayes 250*. 2013.

[178] N. SELLAMI, A. ROUMY et I. FIJALKOW. “A Proof of Convergence of the MAP Turbo-Detector to the AWGN Case”. In : *Signal Processing, IEEE Transactions on* 56.4 (avr. 2008), pages 1548–1561.

[179] K. SLAVAKIS, G.B. GIANNAKIS et G. MATEOS. “Modeling and Optimization for Big Data Analytics : (Statistical) learning tools for our era of data deluge”. In : *IEEE Signal Processing Magazine* 31.5 (sept. 2014), pages 18–31.

[180] D.T.M SLOCK. “Blind fractionaly-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction”. In : *ICASSP*. Avr. 1994.

[181] D.T.M. SLOCK et C.B. PAPADIAS. “Further results on blind identification and equalization of multiple FIR channels”. In : *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*. Tome 3. Mai 1995, 1964–1967 vol.3.

[182] S. SONG et A.C. SINGER. “Pilot-Aided OFDM Channel Estimation in the Presence of the Guard Band”. In : *IEEE Transactions on Communications* 55.8 (août 2007), pages 1459–1465.

[183] J. TELLADO-MOURELO. “Peak to average power reduction for multicarrier modulation”. Thèse de doctorat. USA : Standford University, 1999.

[184] L. TONG, G. XU et T. KAILATH. “A new approach to blind identification and equalization of multipath channels”. In : *Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*. Tome 2. Nov. 1991, pages 856–860.

[185] S. TOUATI et al. “Semi-blind channel estimation for OFDM systems via an EM-Block algorithm”. In : *12th European Signal Processing Conference*. Sept. 2004, pages 2079–2082.

[186] P-Y. TSAI et T-D. CHIUEH. “Frequency-domain interpolation-based channel estimation in pilot-aided OFDM systems”. In : *Vehicular Technology Conference, VTC 2004-Spring.* Tome 1. Mai 2004, 420–424 Vol.1.

[187] M. K. TSATSANIS, G. B. GIANNAKIS et G. ZHOU. “Estimation and equalization of fading channels with random coefficients”. In : *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96*. Tome 2. Atlanta, GA, USA, mai 1996, pages 1093–1096.

[188] M.C. VALENTI et X. XIANG. “Constellation Shaping for Bit-Interleaved Coded APSK”. In : *IEEE International Conference on Communications (ICC)*. Juin 2011, pages 1–5.

[189] M.C. VALENTI et X. XIANG. "Constellation Shaping for Bit-Interleaved LDPC Coded APSK". In : *IEEE Transactions on Communications* 60.10 (oct. 2012), pages 2960–2970.

[190] T.X. VU et al. "Performance analysis of relay networks with channel code in low SNR regime". In : *Signal Processing Advances in Wireless Communications (SPAWC), 2013 IEEE 14th Workshop on*. Juin 2013, pages 575–579.

[191] J.M. WALSH, C.R. JOHNSON et P.A. REGALIA. "A refined information geometric interpretation of turbo decoding". In : *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*. Tome 3. Mar. 2005, pages 481–484.

[192] J.M. WALSH, P.A. REGALIA et C.R. JOHNSON. "Turbo Decoding as Iterative Constrained Maximum-Likelihood Sequence Detection". In : *IEEE Transactions on Information Theory* 52.12 (déc. 2006), pages 5426–5437.

[193] N. WIBERG. "Codes and decoding on general graphs". Thèse de doctorat. Sweden : Likoping University, 1996.

[194] J. WINN, C.M. BISHOP et T. JAAKKOLA. "Variational message passing". In : *Journal of Machine Learning Research* 6 (2005), pages 661–694.

[195] J.K. WOLF. "Redundancy, the Discrete Fourier Transform and impulse noise cancellation". In : *IEEE Trans on Comm.* 31 (mar. 1983), pp.458–461.

[196] J.-L WU et J. SHIU. "Discrete Cosine Transform in error control coding". In : *IEEE Trans on Comm.* 43 (mai 1995), pp.1857–1861.

[197] A. D. WYNER. "The Wire-Tap Channel". In : *Bell System Technical Journal* 54.8 (1975), pages 1355–1387.

[198] G. XU et al. "A least-squares approach to blind channel identification". In : *IEEE Transactions on Signal Processing* 43.12 (déc. 1995), pages 2982–2993.

[199] W. XU et J. ROMME. "A class of multirate convolutional codes by dummy bit insertion". In : *IEEE Global Telecommunications Conference, 2000. GLOBECOM '00.* Tome 2. 2000, 830–834 vol.2.

[200] Y. XU et al. "Variable LLR Scaling in Min-Sum Decoding for Irregular LDPC Codes". In : *Broadcasting, IEEE Transactions on* 60.4 (déc. 2014), pages 606–613.

[201] K. YASUI et T. MATSUSHIMA. "Toward computing the capacity region of degraded broadcast channel". In : *IEEE International Symposium on Information Theory Proceedings (ISIT)*. Juin 2010, pages 570–574.

[202] J. S. YEDIDIA, W.T. FREEMAN et Y. WEISS. "Constructing free-energy approximations and generalized belief propagation algorithms". In : *IEEE Transactions on Information Theory* 51.7 (juil. 2005), pages 2282–2312.

[203] A. ZANKO, A. LESHEM et E. ZEHAVI. "Iterative decoding of robust analog product codes". In : *Electrical & Electronics Engineers in Israel (IEEEI), 2014 IEEE 28th Convention of*. IEEE. 2014, pages 1–5.

[204] J. ZHANG et al. "Two-dimensional correction for min-sum decoding of irregular LDPC codes". In : *IEEE Communications Letters* 10.3 (mar. 2006), pages 180–182.

[205] Q. ZHAO et L. TONG. "Adaptive Blind channel estimation by least squares smoothing". In : *IEEE trans. on Signal Processing* 47 (nov. 1999), pages 3000–3012.

[206] X. ZHUANG et F. VOOK. "Iterative channel estimation and decoding for a turbo OFDM system via the EM algorithm". In : *Proc. ICASSP*. Mai 2002.

Publications

Annexe 1

F. Abdelkefi, P. Duhamel, F. Alberge, Impulsive noise cancellation in multicarrier transmission. IEEE Trans. on Communications, vol 53, n° 1, January 2005.

Abstract

A parallel between Reed–Solomon codes in the complex field and multicarrier transmission using orthogonal frequency-division multiplexing (OFDM) is first presented. This shows that when the signal is sent over some channel composed of Gaussian plus impulsive noise, the impulsive noise can be removed by a procedure similar to channel decoding, using information carried by the "syndrome." This result is first derived in a simple situation (oversampled discrete multitone, additive channel), which is merely of theoretical interest. In any case, consecutive zeros, in the output of the OFDM modulator, do not correspond to real subcarriers. Pilot tones are transmitted for synchronization or channel-estimation purposes. These pilot tones are generally scattered among the information ones. Our approach is to use these pilot tones as syndromes, in order to correct impulsive noise. We show that the correction capacity is conditioned by the position of these pilot tones in the transmitted sequence. A protection subsystem based on hypotheses tests is introduced after the decoding operation in order to detect malfunctions of this decoder. The efficiency of this technique is corroborated with simulations in the slightly modified Hiperlan2 context. Other extensions are then provided in order to increase the practical usefulness of the method.

Impulsive Noise Cancellation in Multicarrier Transmission

Fatma Abdelkefi, Pierre Duhamel, *Fellow, IEEE*, and Florence Alberge

***Abstract*—A parallel between Reed–Solomon codes in the complex field and multicarrier transmission using orthogonal frequency-division multiplexing (OFDM) is first presented. This shows that when the signal is sent over some channel composed of Gaussian plus impulsive noise, the impulsive noise can be removed by a procedure similar to channel decoding, using information carried by the "syndrome." This result is first derived in a simple situation (oversampled discrete multitone, additive channel), which is merely of theoretical interest. In any case, consecutive zeros, in the output of the OFDM modulator, do not correspond to real subcarriers. Pilot tones are transmitted for synchronization or channel-estimation purposes. These pilot tones are generally scattered among the information ones. Our approach is to use these pilot tones as syndromes, in order to correct impulsive noise. We show that the correction capacity is conditioned by the position of these pilot tones in the transmitted sequence. A protection subsystem based on hypotheses tests is introduced after the decoding operation in order to detect malfunctions of this decoder. The efficiency of this technique is corroborated with simulations in the slightly modified Hiperlan2 context. Other extensions are then provided in order to increase the practical usefulness of the method.**

***Index Terms*—Impulsive noise cancellation, multicarrier systems, orthogonal frequency-division multiplexing (OFDM), Reed–Solomon (RS) codes.**

I. Introduction

A MULTICARRIER or orthogonal frequency-division multiplexing (OFDM) system is a method of data modulation that has recently gained increased interest with the development of faster signal-processing components and technologies [1]. It is used in European digital audio broadcasting (DAB) [2], and in wireless environments such as digital broadcast television and mobile communication systems [2], [3]. However, OFDM based on discrete multitone (DMT) systems are also examined for broadband digital communications on the existing copper networks [4]. This technique has been proposed for high-rate and asymmetric digital subcarrier lines (HDSL, ADSL) [4], local area networks (Hiperlan2), etc.

The main idea behind OFDM is to split the transmitted data sequence into N parallel sequences of symbols. This structure has the particularity to enable a simple equalization scheme and to resist to multipath propagation channel. In fact, intersymbol interference (ISI) can be avoided when a guard interval (GI) is implemented between each block of time-domain samples to be transmitted. However, some carriers can be strongly attenuated. It is then necessary to incorporate a powerful channel encoder, combined with frequency and time interleaving. In this way, close-coded bits are not likely to fall simultaneously in a spectral null [2].

However, some of these quoted applications suffer from impulsive noise, and then the performance of such systems is damaged. The impulsive noise is an additive disturbance that arises primarily from the switching electric equipment [5]–[7] and as spectral properties, the defined pulse has a pole and infinite energy. Therefore, bursty or isolated errors are usually generated by an impulsive noise affecting consecutive symbols in the Viterbi decoder, because such a decoder relies on the past history of the symbol sequence. Thus, a powerful decoder is required for such applications, that is robust against impulsive noise in order to minimize its impact. The impulsive noise model that is used in this paper is that given by Ghosh [8], and it will be described later in the paper.

In order to implement a digital modulator, an oversampled version of the continuous signal is often computed. This amounts to appending consecutive null symbols to the block of symbols to be modulated. Therefore, the OFDM modulator can be seen as a real Reed–Solomon (RS) encoder [9], and then it can be used as some specific impulsive noise canceler, the structure of which is well suited to the nature of the problem (i.e., a single impulse shows up as a single error), rather than counting on the classical channel coder to solve the problem. Practically, both types of codes will have to cooperate, in order to process both Gaussian and impulsive noise.

To cancel impulsive noise, Wolf [10], Redinbo [11], Wu [12], and Kumaresan [13] had also used the Bose–Chaudhuri–Hocquengem (BCH) code in the complex or real field, and they have considered the effect of minor errors.

Wolf [10] suggested two methods to correct large errors. The first is based on the Fourier transform coding and is a voting scheme. This technique takes the discrete Fourier transform (DFT) of the received sequence and examines those frequency components that should be zeros for the original data sequence. That means any k samples of the N received samples can be used to estimate the transmitted samples. So there are C_N^k possibilities. If there are small independent errors on each of the transmitted components, the vote of the "correct vector" yields a cluster of vectors rather than a single vector. This technique is impractical for large values of N. The second

Paper approved by C. Tellambura, the Editor for Modulation and Signal Design of the IEEE Communications Society. Manuscript received July 15, 2003.

F. Abdelkefi was with Supelec/LSS, 91192 Gif-sur-Yvette Cedex, France. She is now with the Communication Technology Laboratory, Swiss Federal Institute of Technology, CH-8092 Zurich, Switzerland (e-mail: abdelkefi@nari@ee.ethz.ch).

P. Duhamel is with Supelec/LSS, 91192 Gif-sur-Yvette Cedex, France (e-mail: pierre.duhamel@lss.supelec.fr).

F. Alberge was with Supelec/LSS, 91192 Gif-sur-Yvette Cedex, France. She is now with Orsay University, 91 400 Orsay, France.

Digital Object Identifier 10.1109/TCOMM.2004.840628

method is a slightly modified BCH decoding algorithm, and it is not effective for multiple errors.

Marvasti showed in [14] that the problem of signal reconstruction from missing samples can be resolved by using a reconstruction algorithm similar to the RS decoding technique, based on the Fourier transform. He proposed an error-recovery technique for bursts (BERT) of real and complex samples that uses techniques similar to Peterson's method for BCH decoding and Forney's decoding. The BERT technique is found to be sensitive to background noise.

Wu [12], [15] defined a class of real-number linear block codes using the discrete cosine transform (DCT). Despite the noncyclic nature of the DCT codes, a set of modified syndromes can be defined, with which a modified BCH decoding algorithm can be performed. He supposed that the codeword is corrupted by a minor error vector due to the background noise, and by impulsive noise due to the channel noise. To correct impulsive noise, he used a modified Berlekamp–Massey algorithm and the Forney algorithm, which include a decision threshold. However, for our case, DFT is preferred over DCT due to its cyclic properties.

Kumaresan [13] used RS codes, in real or complex fields to correct bursty impulsive noise in the presence of minor errors in each component of the received vector. He devised several decoding strategies based on least-squares techniques and singular value decomposition to estimate the location and the number of impulsive errors.

However, Kumaresan *et al.* have not suggested any method to control the malfunction of their proposed decoding algorithm. Redinbo presented in [11] and [16] a decoding procedure for real-number codes which are also constructed by imposing constraints in the DFT domain: consecutive zeros. He assumed that codewords are corrupted by small levels of roundoff noise, and occasionally by a few large "impulsive noises." The error-correcting algorithm is divided in two parts. The first is the large activity detection that determines if large excursions are present and estimates their locations. The second part is the large error value estimation. To estimate the impulsive error locations, he used a modified Berlekamp–Massey algorithm. The final stage of this algorithm consists of testing the corrected outputs by recomputing syndromes and employing a threshold detector. Redinbo differs from the others by this protection subsystem, but the proposed threshold is only tuned by simulation.

These quoted references consider that the transmitted sequences contain some consecutive zeros, but this assumption does not correspond to a practical case. In many cases, however, pilot tones are transmitted and are scattered among the information ones. We believe that no studies have been done in order to correct impulsive noise in this context (nonconsecutive zeros or pilots), and by using the properties of BCH codes in the complex or the real field.

To correct impulsive noise in this case, the correction capacity should be defined. However, Hartmann and Tzeng observed that there exist many cyclic codes whose defining set of zeros contains more than one set of consecutive zeros [17], and they succeeded in improving the BCH bound [18]. They extended the BCH bound to that case, but they did not give a general solution. In [9], we have used a special case of the Hartmann–Tzeng bound, by considering that the output of the OFDM modulator contains $(2t)$ uniformly spaced pilot tones, the spacing being coprime with the length of the transmitted sequence. However, some additional flexibility would be very useful in many applications.

To correct impulsive noise in the OFDM system, we have suggested using the similarity between RS codes in the complex field and the OFDM system, and then the properties of RS codes in the complex field are easily applied. Note that the proposed decoding algorithm uses techniques that are practically similar to those used by Wolf [10], Redinbo [11], Wu [12], Maravasti [14], and Kumaresan [13], but we applied these techniques in the general case, where the pilot tones are neither consecutive nor uniformly distributed.

This paper first states the conditions on the locations of the pilot tones so that they can be seen as additional syndromes to correct impulsive noise. This condition is first stated in terms of the rank of some specific matrix. So, we considered the general case, where the Hartmann bound is a particular case.

In the second part of this paper, to detect malfunctions of the decoding algorithm, we propose an *a posteriori* control test which is based essentially on the hypotheses tests. So the threshold that we use is deduced from the receiver operating characteristic (ROC) curve and Bayes criteria.

Implementing a digital modulator requires working with an oversampled version of the transmitted analog signal, and this is equivalent to adding consecutive zeros to the block of symbols to be modulated. If the receiver has the same structure, null symbols should be received at the same locations. We show, in Section II, the similarity between RS codes and the OFDM modulator. Pilot tones can also be used as additional syndromes. In Section III, we establish the conditions on the locations of the pilot tones in order to have a maximal correction capacity. In Section IV, we explain the decoding algorithm. To detect the malfunction of this decoding algorithm, we present in Section V the *a posteriori* control test. Finally, the efficiency of our technique is proved in a slightly modified Hiperlan2 context.

II. Transmission Scheme and Connection With Spectral Codes

A. Transmission Scheme

A discrete model of the OFDM system is easily obtained by computing M samples of the signal to be sent to the channel during one OFDM symbol, i.e., $MT_e = NT_s$, $T_e \leq T_s$ (T_e is the sampling period, and T_s is the OFDM symbol period). Moreover, if one considers the simple multicarrier system where the prototype filter is a rectangular pulse of duration NT_s, modulated with spacing between carriers equal to $1/NT_s$, these samples are computed as

$$c_k(n) = \sum_{m=0}^{N-1} E_m(n-1) e^{\frac{2j\pi km}{M}}$$

which is exactly the inverse discrete Fourier transform (IDFT) of the transmitted sequence $\{E_m(n-1)\}$ enlarged by $(M-N)$ zeros.

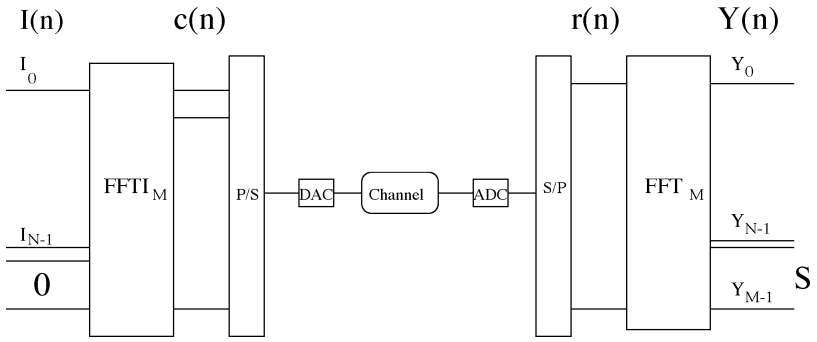


Fig. 1. OFDM system.

At the receiver, the analog-to-digital converter (ADC) samples the signal $r(n)$ at rate T_e, and a DFT is performed. Therefore, the received signal is converted into the frequency domain $\{Y_k\}$

$$Y_k = E_k + N_k, \quad 0 \le k \le M-1$$

where N_k is the Fourier transform of the noise sequence $\{n_k\}$ (see Fig. 1).

This section considers channels without ISI, for simplicity. Extension to the case where the channel introduces ISI is fairly trivial by using a cyclic prefix, as is classically done in OFDM systems.

B. Channel Model

Assuming a memoryless channel, each transmitted sample is modified by the channel according to

$$r_k = c_k + b_k + i_k, \quad k \in \{0 \dots M-1\}$$

where b_k is additive white Gaussian noise (AWGN) with zero mean and variance σ_b^2, and i_k is the impulsive noise. In the following, the impulsive noise is modeled, as in [8], as:

$$i_k = l_k g_k \; \forall k \in \{0 \dots M-1\}$$

where l_k stands for a Bernoulli process, an independent and identically distributed (i.i.d.) sequence of zeros and ones with $\text{prob}(l_k = 1) = p$, and g_k is a complex white Gaussian noise with zero mean and variance σ_i^2, such as $\sigma_i^2 \gg \sigma_b^2$. Note that this model assumes the presence of a large interleaver, so that bursts of errors can be scattered along time.

Under this model, the probability density function (pdf) of the channel noise $n_k = b_k + i_k$ can be expressed as

$$p(x) = (1-p)G\left(x, 0, \sigma_b^2\right) + pG\left(x, 0, \left(\sigma_i^2 + \sigma_b^2\right)\right)$$

where $G(n, m_x, \sigma_x) = (1/\sigma_x\sqrt{2\pi})\exp(-((x-m_x)^2)/2\sigma_x^2)$ (i.e., the Gaussian density with mean m_x and variance σ_x^2).

This expression allows computing the capacity of this channel, in order to estimate the impact of a given impulsive noise on the capacity of a Gaussian channel. An efficient numerical technique to calculate this capacity was derived in 1972 by Blahut and Arimoto [19], [20]. They proposed an iterative procedure which has the property of monotonic convergence to the capacity, and which is applicable to arbitrary discrete memoryless channels. This method has been applied to compute the capacity of the "Gaussian plus Bernoulli Gaussian" channel in the case of the real field. According to the memoryless channel model defined above, the received signal has the expression $r_k = c_k + n_k$, where $\{n_k\}$ stands for the complex noise in the form $n_k = n_{c_k} + jn_{s_k}$, $j = \sqrt{-1}$, and $\{c_k\}$ is the transmitted sequence of the form $c_k = c_{c_k} + jc_{s_k}$. The subscripts "c" and "s" suggest the real and imaginary parts. The expectation of a real random variable is naturally generalized to the complex case as $E(X) = E[X_c] + jE[X_s]$ ($X = X_c + jX_s$, where X is a random variable) [21]. The statistic properties of $X = X_c + jX_s$ are determined by the joint pdf $p_{X_cX_s}(x_c, x_s)$ of X_c and X_s, provided, of course, that the pdf exists: $p_X(x_c + jx_s) = p_{X_cX_s}(x_c, x_s)$. We suppose that $\underline{n}$ and $\underline{c}$ are independent.

$\{n_{c_k}\}$ and $\{n_{s_k}\}$ (respectively, $\{c_{c_k}\}$ and $\{c_{s_k}\}$) have the same autocorrelation function, a vanishing crosscorrelation function, with zero mean and $E[|c_{c_k}|^2] = E[|c_{s_k}|^2]$.

The mutual information between the channel input $\underline{c}$ and the channel output $\underline{r}$ can be written as a function of the entropy $H(\underline{r})$ and the conditional entropy $H(\underline{r}|\underline{c})$. That means

$$I(\underline{c};\, \underline{r}) = H(\underline{r}) - H(\underline{r}|\underline{c}) \tag{1}$$

where $I(.;.)$ stands for the mutual information, H is the entropy, and $\cdot$ denotes vector.

However, one can easily verify that [22]

$$H(\underline{r}|\underline{c}) = H(\underline{n}) \;\text{ and }\; H(\underline{n}) = H(\underline{n_c}) + H(\underline{n_s}). \tag{2}$$

We suppose in the following that real and imaginary parts are independent, thus:

$$H(\underline{r}) = H(\underline{r_c}) + H(\underline{r_s}). \tag{3}$$

From (1)–(3), we deduce that

$$I(\underline{c}\,;\, \underline{r}) = 2I(\underline{c_c}). \tag{4}$$

Thus, $C^{2D} = 2C^{1D}$ where C^{1D} and C^{2D} are, respectively, the capacities of the real channel (which is calculated above) and the complex channel. Fig. 2 depicts the capacity of the "Gaussian plus Bernoulli Gaussian" channel in bits per second normalized by the bandwidth of the channel (W), as a function of the signal power P for several values of p, where $\sigma_i = 1$ and $\sigma_b = 6.10^{-2}$. We note that even for somewhat large values of p, the capacity of the channel is approximately similar to that of the AWGN channel. For example, if $p = 10^{-2}$, and $P = 1$, the capacity of

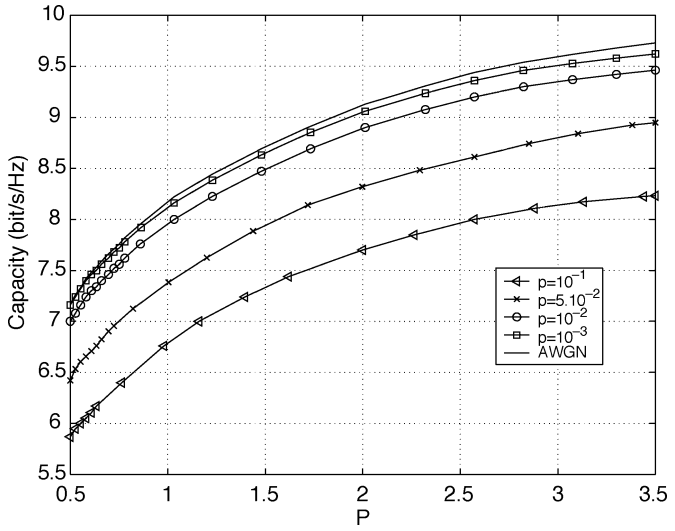


Fig. 2. "Gaussian plus Bernoulli Gaussian" channel capacity.

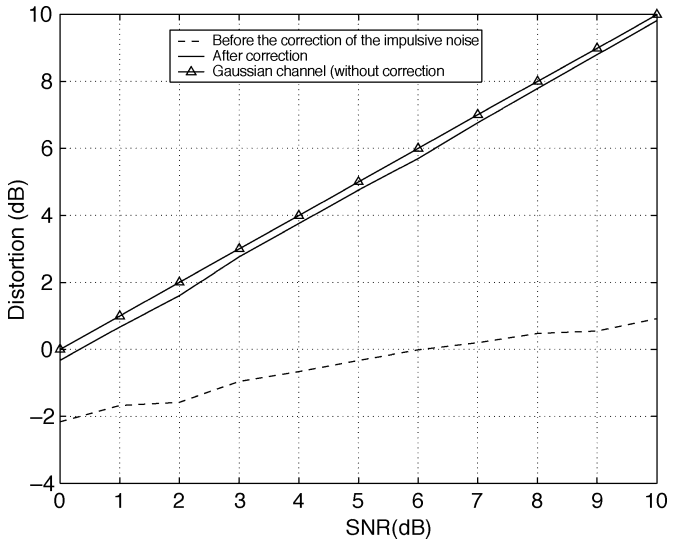


Fig. 3. Distortion performance when we consider a "Gaussian plus Bernoulli Gaussian" channel and consecutive syndrome locations.

the "Gaussian plus Bernoulli Gaussian" channel is 8 b/s/Hz, which is approximately the same value as for the AWGN channel. If $p = 5.10^{-2}$, then we can transmit 7.4 b/s/Hz, which means that we lost only 0.6 b/s/Hz, this decrease of capacity being due to the impulsive noise. The distortion performance of the "Gaussian plus Bernoulli Gaussian" channel is shown in Fig. 3.

However, if no specific procedure is used in an OFDM system, it is unlikely that such similar performance can be obtained. Consider the case of a 64-quadrature amplitude modulation (QAM) constellation transmitted over 64 subbands. Each impulse drastically impairs 384 bits at a time, and it can be stated that the OFDM demodulator acts as an impulsive noise amplifier. This is clearly in favor of a process taking into account the specific nature of the impulsive noise in the OFDM system.

C. Connection Between OFDM System and RS Code

It has been shown in [23]–[25] that the ideas of spectral coding theory can be translated in the frequency domain, over a field $\mathbb{F}$ such that $\mathbb{C}, \mathbb{R}, \ldots$. RS codes can be defined as follows [24].

Definition 1: Let $\mathbb{F}$ contain an element of order M. The $(M, M-2t)$ RS block length M with symbols in $\mathbb{F}$ is the set of all vectors $\underline{c}$ whose spectrum in $\mathbb{F}$ satisfies $C_k = 0, \forall k \in \mathcal{A}$, where $\mathcal{A} = \{k_0 + 1 \ldots k_0 + 2t\}$. This is described briefly as an $(M, M-2t)$ RS code over $\mathbb{F}$.

The spectrum of a RS codeword lives in the same field as information symbols. Then, to form a RS code, a block of $2t$ consecutive spectral components are chosen as parity frequencies (to be set to zero), and the remaining are information symbols. Marshall [26] has shown that a conventional decoding algorithm for finite field cyclic codes could be employed for real and complex numbers. The basic remark that we have used in [9] is that a discrete sequence of complex numbers containing $2t$ consecutive zeros are transmitted over the OFDM system, therefore, the output of the OFDM modulator can be considered as a RS codeword. After transmission over a "Gaussian plus Bernoulli Gaussian" channel, the DFT of the received discrete-time sequence no longer has $2t$ zeros, and this is due only to the channel. Hence, the OFDM modulator can be seen as a complex-valued RS code, and the correction capacity is given by the following BCH bound property.

BCH Bound 1: If ($2t$ consecutive frequencies belong to $\mathcal{A}$) then (the minimal distance is at least $2t-1$), where $\mathcal{A}$ is the set of the $2t$ zeros.

The BCH bound proves that t errors in any codeword of a RS code can always be corrected, because every pair of codewords differs in at least $2t+1$ places. So the correction capacity is up bounded by $\lceil (2t+1)/2 \rceil$.

However, strictly speaking, there are more than $\lceil (2t+1)/2 \rceil$ errors if one uses our channel model; all samples are polluted by noise. Therefore, we concentrate on the removal of the sole impulsive noise, considering the Gaussian component as background noise. The classical decoding techniques have to be adapted to the presence of this background noise.

However, consecutive zeros do not correspond to a part of the OFDM spectrum which is actually available (analog shaping filters limit bandwidth), and only a small part of these zeros can be practically used. In many cases, pilot tones are transmitted for synchronization or channel-estimation purposes. These pilot tones consist of known symbols that are scattered among the information ones. The next paragraph states the conditions on the locations of the pilot tones in order to correct impulsive noise. We believe that this technique is new in the theory of impulsive noise cancellation.

In the following, the corresponding received components of $\{Y_k\}$ will no longer be null (Fig. 1):

$$Y_k = E_k + B_k + I_k \quad \forall k \in \mathcal{A}$$

where I_k is the DFT of the impulsive noise i_n, and B_k that of the background noise b_n. $\mathcal{A}$ is the set of the locations of pilot tones in the transmitted sequence.

Let $\beta = \mathrm{card}\{\mathcal{A}\}$, and $\mathcal{A}(k)$ is the kth element in $\mathcal{A}$. At the receiver, the correction of impulsive noise must operate on the syndromes S_k, which are given by

$$\begin{aligned} S_k &= Y_{\mathcal{A}(k)} - E_{\mathcal{A}(k)}, \quad k \in \{1, \ldots, \beta\} \\ &= B_{\mathcal{A}(k)} + I_{\mathcal{A}(k)} \\ &= \sum_{n=0}^{M-1} b_n W_M^{n\mathcal{A}(k)} + \sum_{m=0}^{\nu-1} i_{f_m} W_M^{f_m \mathcal{A}(k)} \end{aligned} \tag{5}$$

where $W_M = \exp(-j(2\pi/M))$, ν is the number of impulsive noises in the channel, and $\{f_k\}_{k\in\{0,\ldots,\nu-1\}}$ are the locations of impulsive noise in the sequence.

There are two contributions in these terms (6): one is the Fourier transform of the Gaussian background noise, hence, is still Gaussian, and the other one is a sum of Fourier transforms of impulses, hence, is a sum of complex sinusoids, the frequencies of which correspond to the location of the errors. The decoding problem is the estimation of the number of sinusoids, together with their frequencies and amplitudes, polluted by background noise. The two main differences with classical signal-processing situations are the number of samples is orders of magnitude smaller than usual, and one has the knowledge that the frequencies take integer values.

III. Impulsive Noise Localization

In the following, we assume that $\mathcal{A}$ is the set of the location of the pilot tones in the transmitted sequence and $\beta = \mathrm{card}(\mathcal{A})$.

Classically, we define a suppressing sequence $\{\lambda_k\}_{k\in\{0,\ldots,M-1\}}$ (M is the length of the transmitted sequence) as follows:

If $i_l \neq 0$ then $\lambda_l = 0$.

$$\text{That means } \lambda_l i_l = 0 \; \forall l \in \{0, \ldots, M-1\} \tag{6}$$

where $\{i_l\}_{l\in\{0,\ldots,M-1\}}$ is the impulsive noise sequence.

Let $\underline{\underline{\mathcal{F}_M}}$ denote the Fourier matrix of size M. In the frequency domain, (6) reads

$$\underline{\underline{I}}\,\underline{\Lambda} = \underline{0} \tag{7}$$

where $\underline{\underline{I}} = \underline{\underline{\mathcal{F}_M}}(\mathrm{diag}(\underline{i}))\underline{\underline{\mathcal{F}_M^{-1}}}$ and $\underline{\Lambda} = \underline{\underline{\mathcal{F}_M}}\,\underline{\lambda}$.

So the key equation is equivalent to (7) and $\mathrm{rank}(\underline{\underline{I}}) = \mathrm{rank}(\mathrm{diag}(\underline{i}))$, where $\mathrm{diag}(\underline{x})$ denotes a diagonal matrix that contains on the diagonal the components of a vector $\underline{x}$.

A. Solving for the Impulsive Noise Localization

Matrix $\underline{\underline{I}}$ contains known components whose indexes are in $\mathcal{A}$ and the others are not. So the main idea is to group together the maximum of the components of $\underline{\underline{I}}$ whose indexes are in $\mathcal{A}$, in a submatrix $\underline{\underline{I}}^{(r)}$ such that

$$\underline{\underline{I}}^{(r)} = \underline{\underline{H}}\,\underline{\underline{I}}\,\underline{\underline{D}} \tag{8}$$

where $\underline{\underline{H}}$ and $\underline{\underline{D}}$ are selection matrices that depend on $\mathcal{A}$. Note that when the syndromes are consecutive or regularly distributed, it is always possible to find a submatrix containing all the syndromes (that correspond to the components with indexes belonging to $\mathcal{A}$). However, in the general case (randomly distributed syndromes), the task is much more difficult. Thus the size of the matrix $\underline{\underline{I}}^{(r)}$ puts a limit on the correction capacity. Let the size of $\underline{\underline{H}}$ be $((r+s)\times M)$, and that of $\underline{\underline{D}}$ be $(M \times (r+1))$ where r and s are positive integers. Therefore, we have rank $(\underline{\underline{I}}^{(r)}) \leqslant r+1$, and then we can correct at most r errors. So we suppose that we have at most r impulsive noise (since we can not correct more than r impulsive noise), then there are $M-r$ degrees of freedom (DOF) on $\{\lambda_k\}$, and this is also equivalent to saying that there are $M-r$ DOF on $\{\Lambda_k\}$ (because $\underline{\Lambda} = \underline{\underline{\mathcal{F}_M}}\,\underline{\lambda}$). As the size of the selected matrix $\underline{\underline{I}}^{(r)}$ is $((r+s)\times(r+1))$ and the DOF on $\{\Lambda_k\}$ is $M-r$, then it is possible to choose $M-r-1$ values of $\{\Lambda_k\}$ equal to zero (note that this is equivalent to selecting from $\underline{\underline{I}}$ the matrix $\underline{\underline{I}}^{(r)}$) and one value equal to one. This is equivalent to multiplying $\underline{\Lambda}$ by $\underline{\underline{D}}^t$. Let $\underline{\Lambda}^{(r)} = (\Lambda_0^r, \ldots, \Lambda_r^r)^t = \underline{\underline{D}}^t\underline{\Lambda}$ and $\Lambda_0^{(r)} = 1$, then the key equation (7) is reduced to

$$\underline{\underline{I}}^{(r)}\underline{\Lambda}^{(r)} = \underline{0}. \tag{9}$$

If rank $(\underline{\underline{I}}^{(r)}) = r$, then the key equation (9) has a single solution. Therefore, an important question for being able to solve (9) is: Under which conditions is this system full rank?

If $\underline{\underline{I}}^{(r)}$ has a full row rank, then the correction capacity is r. By construction, $\underline{\underline{I}}^{(r)}$ has the following structure:

$$\begin{pmatrix} I_{m_0+\theta_0+\delta_0} & I_{m_0+\theta_0+\delta_1} & \cdots & I_{m_0+\theta_0+\delta_{r-1}} \\ I_{m_0+\theta_1+\delta_0} & I_{m_0+\theta_1+\delta_1} & \cdots & I_{m_0+\theta_1+\delta_{r-1}} \\ \vdots & \vdots & \ddots & \vdots \\ I_{m_0+\theta_{r+s-1}+\delta_0} & I_{m_0+\theta_{r+s-1}+\delta_1} & \cdots & I_{m_0+\theta_{r+s-1}+\delta_{r-1}} \end{pmatrix} \tag{10}$$

where m_0, $(\theta_k)_{0\leq k\leq r+s-1}$, and $(\delta_k)_{0\leq k\leq r-1}$ are integers such that $0 \leq \theta_k \leq M-1$ and $0 \leq \delta_k \leq M-1$, and that depend on the syndrome location in the sequence.

It has been seen in [27] that the submatrix $\underline{\underline{I}}_r^{(r)}$ constructed with the r first columns and rows of the matrix $\underline{\underline{I}}^{(r)}$ can be decomposed as $\underline{\underline{I}}_r^{(r)} = \underline{\underline{\mathcal{Q}}}^{(r)}\,\underline{\underline{\mathcal{P}}}^{(r)}\,\underline{\underline{\mathcal{R}}}^{(r)}$, where

$$\underline{\underline{\mathcal{Q}}}^{(r)} = \begin{pmatrix} W_M^{f_0\theta_0} & W_M^{f_1\theta_0} & \cdots & W_M^{f_{r-1}\theta_0} \\ W_M^{f_0\theta_1} & W_M^{f_1\theta_1} & \cdots & W_M^{f_{r-1}\theta_1} \\ \vdots & \vdots & \ddots & \vdots \\ W_M^{f_0\theta_{r-1}} & W_M^{f_1\theta_{r-1}} & \cdots & W_M^{f_{r-1}\theta_{r-1}} \end{pmatrix} \tag{11}$$

$$\underline{\underline{\mathcal{R}}}^{(r)} = \begin{pmatrix} W_M^{f_0\delta_0} & W_M^{f_0\delta_1} & \cdots & W_M^{f_0\delta_{r-1}} & \\ W_M^{f_1\delta_0} & W_M^{f_1\delta_1} & \cdots & W_M^{f_1\delta_{r-1}} & \\ \vdots & \vdots & \ddots & & \vdots \\ W_M^{f_{r-1}\delta_0} & W_M^{f_{r-1}\delta_1} & \cdots & W_M^{f_{r-1}\delta_{r-1}} & \end{pmatrix} \tag{12}$$

$$\underline{\underline{\mathcal{P}}}^{(r)} = \begin{pmatrix} i_{f_0} W_M^{m_0 f_0} & 0 & \cdots & 0 \\ 0 & i_{f_1} W_M^{m_0 f_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & i_{f_{r-1}} W_M^{m_0 f_{r-1}} \end{pmatrix} \tag{13}$$

where $f_0 \ldots f_{r-1}$ are the locations of the impulses in the sequence such that $0 \leq f_0 < f_1 \ldots < f_{r-1} < M-1$ and $W_M = \exp(-(2\pi j/M))$.

If $\underline{\underline{\mathcal{R}}}^{(r)}$ and $\underline{\underline{\mathcal{Q}}}^{(r)}$ are full rank, then rank $(\underline{\underline{\mathcal{Q}}}^{(r)}\,\underline{\underline{\mathcal{P}}}^{(r)}\,\underline{\underline{\mathcal{R}}}^{(r)}) = r$, thus the key equation (9) has a unique solution, the correction capacity is maximal, and it is at most r.

Since $\underline{\underline{\mathcal{R}}}^{(r)}$ and $\underline{\underline{\mathcal{Q}}}^{(r)}$ have the same structure, we only study the conditions on $\{\theta_k\}$ that ensure that the matrix $\underline{\underline{\mathcal{Q}}}^{(r)}$ is invertible for any error location. These conditions will apply similarly on $\underline{\underline{\mathcal{R}}}^{(r)}$.

Note that Hartmann provides a solution to this problem.

Hartmann Theorem 1: Let $g(x) \in GF(q)[x]$ be the generator polynomial of a cyclic code, V_M, of length M. If $g(\beta^{m_0+k\theta+l\delta}) = 0$ for $k = 0, 1, \ldots d_0 - 2$ and $l = 0, 1, \ldots s$ where $gcd(\theta, M) = 1$ and $gcd(\delta, M) = 1$, then $d_0 + s \leq d$, where d is the minimal distance.

This suggests that the consecutive d_0+s-1 spectral zeros of the BCH bound can be replaced by a pattern of $s+1$ uniformly subblocks, each of $d_0 + s - 1$ uniformly spaced spectral zeros.

However, we would like to search for more general conditions, leading to greatest flexibility for placing the pilot tones. We have developed some necessary conditions on the matrices $\underline{\underline{\mathcal{R}}}^{(r)}$ and $\underline{\underline{\mathcal{Q}}}^{(r)}$ to be invertible in the general case (for all impulsive noise locations), and we have not yet verified that these conditions are also sufficient. This condition is as follows.

Necessary Condition 1: Let $\theta_k^{(r-1)} = \theta_{k+1}^{(r)} - \theta_{k_0}^{(r)} = \theta_{k+1} - \theta_{k_0}$ $\forall k \neq k_0$, where k, $k_0 \in \{0, \ldots, r-1\}$. So, $\underline{\underline{\mathcal{Q}}}^{(r)}$ is not invertible if one of the two following necessary conditions is verified:

1) there is at least one $k_j \neq k_0$ such that $gcd(\theta_{k_j}^{(r-1)}, M) = 1$;
2) there are at most $(r-2)$ values of $\theta_k^{(r-1)}$ that are not equal to $\theta_{k_j}^{(r-1)}$ and that verify $1 < gcd(\theta_k^{(r-1)}, M) \leqslant r-1$.

When $r = 2$ and $r = 3$, we find necessary and sufficient conditions (section below).

B. Locator Polynomial Evaluation

We have seen in the paragraph above that the correction capacity is maximum for a given value of r only if $\underline{\underline{\mathcal{R}}}^{(r)}$ and $\underline{\underline{\mathcal{Q}}}^{(r)}$ are full rank.

According to the key equation (7), we see that the impulsive noise locations are roots of the locator polynomial $\Lambda(x)$, whose coefficients are already calculated by the key equation (9), such that

$$\Lambda\left(W_M^{-f_k}\right) = 0, \quad \forall k \in \{0, 1, \ldots r-1\}. \tag{14}$$

Since we propose that the number of impulsive noise is r, then $M - r - 1$ values of $\{\Lambda_k\}$ are chosen zeros, and this is due to the DOF that we have on $\{\Lambda_k\}$. However, the degree of $\Lambda(x)$ is generally much larger than r (it is equal to r in the contiguous case), since the pilot tones are irregularly spaced. Thus, it has to be checked that no other root of this polynomial will generate a false alarm (zero on the unit circle, at an integer location) which corresponds to $\text{card}\{f_k | \Lambda(W_M^{-f_k}) = 0\} = r$.

To annul $(M - r - 1)$ values of $\underline{\Lambda}$ (in order to cancel the unknown components of $\underline{I}$) is equivalent to multiplying $\underline{\Lambda}$ by the matrix $\underline{\underline{D}}$ [see (8)] and then $\underline{\Lambda} = \underline{\underline{D}}\,\underline{\Lambda}^r$. Thus, we have

$$\underline{\lambda} = \underline{\underline{\mathcal{F}}}_M^{-1}\,\underline{\underline{D}}\,\underline{\Lambda}^{(r)}. \tag{15}$$

So, we verify that

$$\underline{\underline{\mathcal{F}}}_M^{-1}\,\underline{\underline{D}} = \begin{pmatrix} 1 & \cdots & 1 \\ W_M^{\delta_0} & \cdots & W_M^{\delta_r} \\ \vdots & \ddots & \vdots \\ W_M^{M\delta_0} & \cdots & W_M^{M\delta_r} \end{pmatrix}. \tag{16}$$

However, if each $(r+1)$ square submatrix of $\underline{\underline{\mathcal{F}}}_M^{-1}\underline{\underline{D}}$ is invertible, then $\underline{\lambda}$ has at most r zeros, because size$(\underline{\Lambda}^{(r)}) = r + 1$. It is easily seen that this problem is of the same nature as the one in the previous subsection, with a dimension augmented by 1. Thus $\text{card}\{f_k | \Lambda(W_M^{-f_k}) = 0\} = r$ only if $\underline{\underline{\mathcal{Q}}}^{(r+1)}$ is invertible.

C. Examples

This general problem seems intricate, and we are still working on it. However, when $r = 2$ and $r = 3$, we have found that the necessary conditions on $\{\theta_k\}$ and $\{\delta_k\}$ are also sufficient.

If $r = 2$, it can be shown that $\underline{\underline{I}}^{(2)}$ is invertible if

$$gcd(\theta_1 - \theta_0, M) = 1 \text{ and } gcd(\delta_1 - \delta_0, M) = 1. \tag{17}$$

Thus rank $(\underline{\underline{I}}^{(2)}) = 2$. That means that we can correct at most two impulsive noise.

If $r = 3$, define the function f as

$$f(\theta_1 - \theta_0, \theta_2 - \theta_0) = \left|\det\left(\underline{\underline{\mathcal{Q}}}^{(r)}\right)\right|, \quad \forall(\theta_0,\, \theta_1,\, \theta_2) \in \{0, 1, \ldots M-1\}^3.$$

It is easily seen that

$$\begin{aligned} f(\theta_1 - \theta_0, \theta_2 - \theta_0) &= f(\theta_2 - \theta_0, \theta_1 - \theta_0) \\ &= f(\theta_2 - \theta_0, \theta_2 - \theta_1). \end{aligned}$$

As a consequence of these equalities, we prove that the correction capacity is at most three, if and only if one of these following conditions is met:

$$\text{C1}: \begin{cases} gcd(\theta_2 - \theta_0, M) = 1 \\ gcd(\theta_2 - \theta_1, M) = 1 \\ gcd(\theta_1 - \theta_0, M) = \gamma \end{cases}$$

$$\text{C2}: \begin{cases} gcd(\theta_1 - \theta_0, M) = 1 \\ gcd(\theta_2 - \theta_0, M) = 1 \\ gcd(\theta_2 - \theta_1, M) = \gamma \end{cases}$$

$$\text{C3}: \begin{cases} gcd(\theta_2 - \theta_1, M) = 1 \\ gcd(\theta_1 - \theta_0, M) = 1 \\ gcd(\theta_2 - \theta_0, M) = \gamma \end{cases}$$

where $\gamma = 1$ if M is prime, otherwise $\gamma = 2$. We have also the same conditions on $\{\delta_k\}$.

So if $\theta_2 - \theta_1 = \theta_1 - \theta_0$ and if $gcd(\theta_2 - \theta_1, M) = 1$, then the considered matrix is invertible, we verify that this particular case is the Hartmann bound. Thus, we can say that the condition that we find is more general than that of Hartmann.

These two cases ($r = 2$ and $r = 3$) and the general condition (which we have verified is necessary) can be useful in many applications, such as in the practical context of Hiperlan2, where it is possible to cancel impulsive noise [28]–[30] and reduce the peak average power rate (PAPR) level.

IV. Decoding Algorithm

In the proposed procedure, we adapt a classical decoding algorithm to the presence of background noise.

The decoding algorithm works in three steps: estimate the number ν of impulsive errors; find the error locations; and correct the errors. There are several efficient algorithms for doing this, but these are very sensitive to small levels of noise. Therefore, we have used a modified version of the Peterson–Gorenstein–Zierler algorithm adapted to complex field [31], [32] and which is less sensitive to errors. Our work was performed simultaneously in the context of joint source and channel coding [32]. In the following, we have considered the general case. We have supposed that we have in the transmitted sequence some pilot tones which are not consecutive and are known (not zero).

A. Estimation of ν

According to Section III, the syndrome matrix is expressed as shown in the equation at the bottom of the page, where m_0, $(\theta_k)_{0\leq k\leq r+s-1}$, and $(\delta_k)_{0\leq k\leq r-1}$ have already been defined in Section III, and we suppose that this syndrome matrix is full rank.

Let ν denotes the number of impulsive noise such that $\nu < r$.

To remove the contribution of the background noise, we calculate the correlation matrix $\underline{\underline{S}}^{(r)^H}\underline{\underline{S}}^{(r)}$. Since $\underline{\underline{B}}^{(r)}$ and $\underline{\underline{I}}^{(r)}$ are assumed to be uncorrelated with zero means $E[\underline{\underline{B}}^{(r)^H}\ \underline{\underline{I}}^{(r)}] = \underline{\underline{0}}_r$, and since the background noise is supposed to be a white Gaussian noise with variance σ_b^2 and with zero mean, then $E[\underline{\underline{B}}^{(r)^H}\ \underline{\underline{B}}^{(r)}] = r\ \sigma_b^2\ \underline{\underline{Id}}_r$, where $\underline{\underline{Id}}_r$ is the identity matrix of dimension r.

For a great number of observations, and as the amplitude of the impulsive noise is more important (directly connected to the eigenvalues of the matrix $\underline{\underline{\Gamma}}^{(r)^H}\underline{\underline{\Gamma}}^{(r)}$) than that of the background noise, the number of errors is then estimated as the number of eigenvalues of the matrix $\underline{\underline{S}}^{(r)^H}\underline{\underline{S}}^{(r)}$ whose amplitude is superior to $r\sigma_b^2$ [33]. Regrettably, the relatively reduced number of observations does not allow correctly estimating these errors; a multiplier empirical factor ϕ is applied. We find by simulation that $\phi \in [1\ 4]$, and then

$$\underline{\underline{S}}^{(r)^H}\ \underline{\underline{S}}^{(r)} \approx \underline{\underline{I}}^{(r)^H}\ \underline{\underline{I}}^{(r)} + r\phi\sigma_b^2\underline{\underline{Id}}_r.$$

This estimation is almost accurate, hence, $\tilde{\nu}$ can be estimated as the number of eigenvalues of $\underline{\underline{S}}^{(r)^H}\ \underline{\underline{S}}^{(r)}$ greater than $r\phi\sigma_b^2$.

B. Error Localization

According to (9), and as we have $\tilde{\nu}$ impulsive noise, then we have the following system:

$$\begin{bmatrix} I_{m_0+\theta_0+\delta_0} & \cdots & I_{m_0+\theta_0+\delta_{\tilde{\nu}-1}} \\ \vdots & \vdots & \vdots \\ I_{m_0+\theta_{r+s-1}+\delta_0} & \cdots & I_{m_0+\theta_{r+s-1}+\delta_{\tilde{\nu}-1}} \end{bmatrix} \underline{\Lambda}^{(\tilde{\nu})} = -\begin{bmatrix} I_{m_0+\theta_0+\delta_{\tilde{\nu}}} \\ \vdots \\ I_{m_0+\theta_{r+s-1}+\delta_{\tilde{\nu}}} \end{bmatrix}$$

$$\underline{\underline{I}}^{(\tilde{\nu})}\underline{\Lambda}^{(\tilde{\nu})} = \underline{I}^{(\tilde{\nu})}.$$

Taking into account the statistical contribution of a small level of noise, we get

$$\left(\underline{\underline{I}}^{(\tilde{\nu})^H}\underline{\underline{I}}^{(\tilde{\nu})}\right)^{-1} \approx \left(\underline{\underline{S}}^{(\tilde{\nu})^H}\underline{\underline{S}}^{(\tilde{\nu})} - \tilde{\nu}\sigma_b^2\underline{\underline{Id}}_{\tilde{\nu}}\right)^{-1}.$$

As $\underline{\underline{\Gamma}}^{(\tilde{\nu})}$ and $\underline{\underline{I}}^{(\tilde{\nu})}$ are uncorrelated, a good estimation of $\underline{\Lambda}^{(\tilde{\nu})}$ is

$$\underline{\tilde{\Lambda}}^{(\tilde{\nu})} = \left(\underline{\underline{S}}^{(\tilde{\nu})^H}\underline{\underline{S}}^{(\tilde{\nu})} - \tilde{\nu}\sigma_b^2\underline{\underline{Id}}_{\tilde{\nu}}\right)^{-1}\underline{\underline{S}}^{(\tilde{\nu})^H}\underline{S} \tag{18}$$

$$\approx \left(\underline{\underline{I}}^{(\tilde{\nu})^H}\underline{\underline{I}}^{(\tilde{\nu})}\right)^{-1}\underline{\underline{I}}^{(\tilde{\nu})^H}\underline{I}^{(\tilde{\nu})} = \underline{\Lambda}^{(\tilde{\nu})}. \tag{19}$$

So, the $\tilde{\nu}$ error locations are estimated as the $\tilde{\nu}$ indexes $[0 \ldots M-1]$ that give the smallest values of $|\tilde{\Lambda}^{(\tilde{\nu})}(x)|$ taken on $W_M^{-[0\ \cdots\ M-1]}$.

C. Error Amplitude

The expression of the syndrome is $\underline{S} = \underline{\underline{V}}\ \underline{i} + \underline{\underline{W}}\ \underline{b}$, where the matrix $\underline{\underline{V}}$ depends on the impulsive noise locations. $\underline{i}$ is the vector containing the corresponding amplitudes such that $\underline{\underline{V}}(k,j) = W_M^{\mathcal{A}(k)j}, k \in \{1,\ldots,\beta\}$ and $j \in \{0,\ldots,\nu-1\}$, and $\underline{\underline{W}}(k,t) = W_M^{\mathcal{A}(k)t}, k \in \{1,\ldots,\beta\}$ and $t \in \{0,\ldots,M-1\}$.

An estimate $\underline{\tilde{i}}$ of $\underline{i}$ can be obtained by solving the system above in the least-square sense. Then $\underline{\tilde{i}}$ is

$$\underline{\tilde{i}} = (\underline{\underline{\tilde{V}}}^H\ \underline{\underline{\tilde{V}}})^{-1}\ \underline{\underline{\tilde{V}}}^H\ \underline{S}.$$

V. The *A Posteriori* Control

In the last paragraph, we have shown that the decoding algorithm is in three steps: estimate the number ν of impulsive noise; seek the error locations; and correct the impulsive errors. We add a control step which is able to carefully check whether the decoding procedure has worked correctly. In this way, we are able to begin a truncated enumeration of all possible error locations (the most sensitive part of the algorithm) among the

$$\underline{\underline{S}} = \underline{\underline{S}}^{(r)} = \begin{pmatrix} S_{m_0+\theta_0+\delta_0} & S_{m_0+\theta_0+\delta_1} & \cdots & S_{m_0+\theta_0+\delta_{r-1}} \\ S_{m_0+\theta_1+\delta_0} & S_{m_0+\theta_1+\delta_1} & \cdots & S_{m_0+\theta_1+\delta_{r-1}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m_0+\theta_{r+s-1}+\delta_0} & S_{m_0+\theta_{r+s-1}+\delta_1} & \cdots & S_{m_0+\theta_{r+s-1}+\delta_{r-1}} \end{pmatrix} = \underline{\underline{B}}^{(r)} + \underline{\underline{I}}^{(r)}$$

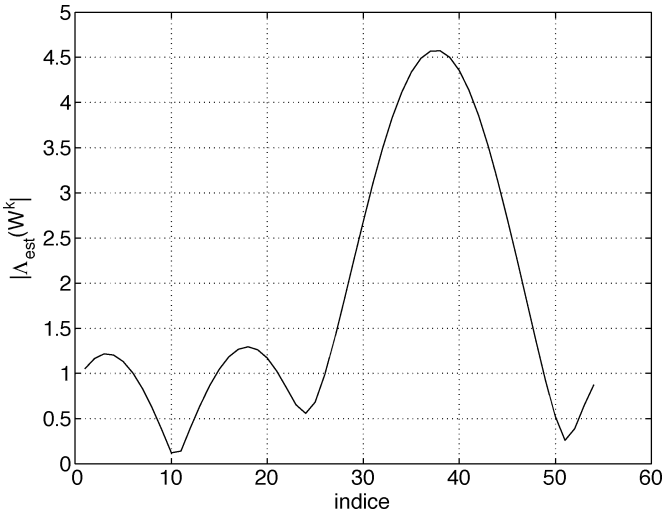


Fig. 4. Error-locator polynomial amplitude.

most likely ones. This truncated enumeration is necessary because of the presence of the background noise which introduces some fuzziness.

Malfunctions of the decoder can be due to:

- wrong estimation of impulsive noise locations. The error locations are linked with the zeros of the location polynomial $\Lambda(W_M)$ (indeed, the zeros of Λ are of the form $W_M^k = \exp(-2j\pi k/M)$). If $\Lambda(W_M^{-k}) = 0$, then k is an impulsive noise location. But here the syndromes contain the Gaussian and Bernoulli noise, thus $\Lambda(W_M^{-k})$ will not be zero and will take small magnitude. Therefore, the localization can be wrong, as shown in Fig. 4, where we plot $|\Lambda(W_M^{-k})|$ versus the position k with $k \in \{0 \dots M-1\}$, $M = 54$, and $\nu = 3$. We observe that at the position near 10, $\Lambda(W_M^{-k})$ is close to zero. As a consequence, the decoding algorithm cannot detect the correct impulsive noise locations;
- wrong estimation of the number of errors;
- possible overflow of error capacity.

Thus, a protection subsystem is introduced after the decoding operation in order to detect malfunctions at each step of the decoding algorithm. In Section IV, we have seen that the expression of the syndrome vector is

$$\underline{S} = \underline{\underline{V}}\,\underline{i} + \underline{\underline{W}}\,\underline{b}$$

where the matrix $\underline{\underline{V}}$ depends on the impulsive noise locations, and $\underline{i}$ is the vector containing the corresponding amplitudes.

The corrected outputs are tested by comparing the syndrome vector with its estimate $\underline{\tilde{S}} = \underline{\underline{\tilde{V}}}\,\underline{\tilde{i}}$ where $\underline{\tilde{i}} = (\underline{\underline{\tilde{V}}}\,\underline{\underline{\tilde{V}}}^H)^{-1}\,\underline{\underline{\tilde{V}}}^H\,\underline{S}$ and $\underline{\underline{\tilde{V}}}$ depends on the estimated locations of the impulsive noise. We calculate $y = \|\underline{\tilde{S}} - \underline{S}\|^2$. If this amount exceeds a certain threshold, then we try to correct malfunctions, else we can conclude that we have properly corrected impulsive noise.

The optimal value of this threshold is obtained thanks to the hypotheses testing theory, as explained below.

In the problem under consideration, either the localization of impulsive noise is correct, which means $Im(\underline{\underline{V}}) \subset Im(\underline{\underline{\tilde{V}}})$, or we have a wrong localization of impulsive noise, which means $Im(\underline{\underline{V}}) \not\subset Im(\underline{\underline{\tilde{V}}})$. We denote by Im the image of the considered space.

1) **If $\mathbf{Im}(\underline{\underline{V}}) \subset \mathbf{Im}(\underline{\underline{\tilde{V}}})$, then** $\underline{\tilde{S}} - \underline{S} = \underline{\underline{P}}\,(\underline{\underline{V}}\,\underline{i} + \underline{\underline{W}}\,\underline{b})$, where $\underline{\underline{P}} = \underline{\underline{Id}}_\beta - \underline{\underline{\tilde{V}}}(\underline{\underline{\tilde{V}}}^H\underline{\underline{\tilde{V}}})^{-1}\,\underline{\underline{\tilde{V}}}^H$ is a projection matrix on $Ker(\underline{\underline{\tilde{V}}})$ and the rank of $\underline{\underline{P}}$ is $(\beta - \tilde{\nu})$, where $\tilde{\nu}$ is the estimated impulsive noise number and $\beta = \mathrm{card}(\mathcal{A})$. Since $Im(\underline{\underline{V}}) \subset Im(\underline{\underline{\tilde{V}}})$, we verify that $\underline{\underline{P}}\,\underline{\underline{V}} = \underline{0}$. So we conclude that

$$\text{If } Im(\underline{\underline{V}}) \subset Im(\underline{\underline{\tilde{V}}}) \text{ then } \|\underline{\tilde{S}} - \underline{S}\|^2 = \|\underline{\underline{P}}\,\underline{\underline{W}}\,\underline{b}\|^2.$$

As the vector $\underline{b}$ is a Gaussian noise of variance σ_b^2 and zero mean, then $\underline{\underline{W}}\underline{b}$ is a Gaussian noise vector that contains $2M$ independent real Gaussian noise such that each one is of variance $\sigma_b^2/2$ and zero mean (2 is due to the fact that we consider the complex field $\mathbb{C}$). And as $\underline{\underline{P}}$ is a projection matrix, thus $\|\underline{\tilde{S}} - \underline{S}\|^2 = \|\underline{\underline{P}}\,\underline{\underline{W}}\underline{b}\|^2$ is a chi-square distribution $(\|\underline{\underline{P}}\,\underline{\underline{W}}\underline{b}\|^2/(\sigma_b^2/2)) \sim \chi^2(2(\beta - \tilde{\nu}))$.

2) **If $\mathbf{Im}(\underline{\underline{V}}) \not\subset \mathbf{Im}(\underline{\underline{\tilde{V}}})$,** we can write $\underline{\tilde{S}} - \underline{S}$ as follows:

$$\underline{\tilde{S}} - \underline{S} = \underline{\underline{P}}[\underline{\underline{W}}\,\underline{\underline{V}}]\begin{pmatrix}\underline{b}\\ \underline{i}\end{pmatrix}.$$

The Gaussian noise and the Bernoulli one do not have the same variances, so we multiply the vector $\binom{\underline{b}}{\underline{i}}$ by a diagonal matrix $\underline{\underline{Q}}$ in order to normalize it. Then the vector $\underline{R} = \underline{\underline{Q}}\binom{\underline{b}}{\underline{i}}$ is a normal Gaussian noise, where $\underline{\underline{Q}} = \mathrm{diag}(((1/\sigma_b)\dots(1/\sigma_b),(1/\sigma_i)\dots(1/\sigma_i)))$.

Let $\underline{\underline{Q}} = \underline{\underline{P}}\,[\underline{\underline{W}}\;\underline{\underline{V}}]\,\underline{\underline{D}}^{-1}$, then $\|\underline{\tilde{S}} - \underline{S}\|^2 = \underline{R}^H\,\underline{\underline{Q}}^H\,\underline{\underline{Q}}\,\underline{R}$. $\underline{\underline{Q}}^H\underline{\underline{Q}}$ is a positive definite, Hermitian matrix, therefore it is diagonalizable, i.e., there exists a unitary matrix $\underline{\underline{K}}$ such that $\underline{\underline{Q}}^H\underline{\underline{Q}} = \underline{\underline{K}}^H\,\underline{\underline{G}}\,\underline{\underline{K}}$, where $\underline{\underline{G}}$ is a diagonal matrix that contains the eigenvalue of the Hermitian matrix. Then, $y = \|\underline{\tilde{S}} - \underline{S}\|^2 = (\underline{\underline{K}}\,\underline{R})^H\,\underline{\underline{G}}\,\underline{\underline{K}}\,\underline{R}$. As $\underline{R}$ is a normal Gaussian vector, then $\underline{Z} = \underline{\underline{K}}\,\underline{R}$ is also a normal Gaussian one. Thus

$$\|\underline{\tilde{S}} - \underline{S}\|^2 = \sum_{k=1}^{\mathrm{rank}(\underline{\underline{Q}})} (g_k Z_k)^2 \tag{20}$$

where (g_k) are the eigenvalues of the matrix $\underline{\underline{Q}}^H\underline{\underline{Q}}$. We can easily verify that $\mathrm{rank}\,(\underline{\underline{Q}}) = \mathrm{rank}(\underline{\underline{P}}) = \beta - \tilde{\nu}$. Thus, $\|\underline{\tilde{S}} - \underline{S}\|^2$ is a linear combination of chi-square distributions of the random variables (Z_k), such that $((g_k Z_k)^2/(g_k^2/2)) \sim \chi^2(2)$.

So, we can deduce the pdf of $y = \|\underline{\tilde{S}} - \underline{S}\|^2$. As g_k depends on the unknown impulsive noise locations, then we calculate the conditional pdfs $p_Y(y|\underline{\mathrm{loc}})$ for each value of the location of the impulsive noise, where $\underline{\mathrm{loc}}$ is the vector containing the locations of the impulsive noise, and we compute the average with respect to the locations. So the Bayes formula leads to this equality

$$p_Y(y) = \sum_{\underline{\mathrm{loc}}} p_Y(y|\underline{\mathrm{loc}})\,p(\underline{\mathrm{Loc}} = \underline{\mathrm{loc}}). \tag{21}$$

Generally, the analytic expression of $p_Y(y)$ is unknown, but it is always possible to calculate it when the correction capacity is definite.

At this point, we have obtained the two pdf expressions that correspond to the two situations under study (i.e., the error localizations are all correct $((\|\tilde{\underline{S}} - \underline{S}\|^2/(\sigma_b^2/2)) \sim \chi^2(2(\beta - \tilde{\nu})))$, or there is at least a wrong localization. The discussion below focuses on the optimal choice for a decision threshold using hypotheses tests. Indeed, we have a decision problem with two hypotheses: 1) H_0: there are not impulsive noises, or there are and we have well localized them; and 2) H_1: we have not well-localized impulsive noise.

We know that either H_0 or H_1 is true. Thus each time the experiment is conducted, one of four things can happen: 1) H_0 is true, choose H_0; 2) H_0 is true, choose H_1; 3) H_1 is true, choose H_1; and 4) H_1 is true, choose H_0.

The first and third alternatives correspond to correct choices. The second and fourth alternatives correspond to errors. Since we assume that the decision rule must say either H_0 or H_1, we can view it as a rule for dividing the total observation space into two parts, Σ_0 and Σ_1. Whenever an observation falls in Σ_0, we say H_0, and whenever an observation falls in Σ_1, we say H_1.

Let $P_F = \int_{\Sigma_1} p_{y|H_0}(Y|H_0)dy$ be the probability of false alarm, and $P_D = \int_{\Sigma_1} p_{y|H_1}(Y|H_1)dy$ the probability of detection. In general, we would like to make P_F as small as possible, and, at the same time, to have P_D as large as possible. Let $\tilde{\underline{\text{loc}}}$ be the estimate of the impulsive noise location vector $(\text{size}(\tilde{\underline{\text{loc}}}) = \tilde{\nu})$.

To define Σ_0, we have to look at two cases:

I $\quad \underline{\text{loc}} = \tilde{\underline{\text{loc}}}$;

II $\quad \underline{\text{loc}} \subset \tilde{\underline{\text{loc}}}$ and $\text{size}(\underline{\text{loc}}) < \text{size}(\tilde{\underline{\text{loc}}})$.

If $\underline{\text{loc}} \subset \tilde{\underline{\text{loc}}}$, then we can easily verify that $Im(\underline{\underline{V}}) \subset Im(\tilde{\underline{\underline{V}}})$ and $\underline{\underline{P}}\,\underline{\underline{V}} = \underline{\underline{P}}\,\tilde{\underline{\underline{V}}} = \underline{\underline{0}}$. That means these two cases are equivalent, and $p_{y|H_0} \sim \chi^2(2(\beta - \tilde{\nu}))$.

For Σ_1, we look at these cases:

I $\quad \underline{\text{loc}} \not\subset \tilde{\underline{\text{loc}}}$ and $\underline{\text{loc}} \bigcap \tilde{\underline{\text{loc}}} = \emptyset$;

II $\quad \underline{\text{loc}} \not\subset \tilde{\underline{\text{loc}}}$ and $\underline{\text{loc}} \bigcap \tilde{\underline{\text{loc}}} \neq \emptyset$.

When P_D and P_F are calculated, we plot P_D versus P_F for various values of impulsive-noise-to-Gaussian-noise ratio (INR) as a parameter on the curve which is often referred to as the receiver operating characteristic (ROC). It completely describes the performance of the test as a function of the parameter of interest. Now, we have to look for the optimal threshold $\delta^{(\tilde{\nu})} = y_{\text{opt}}$, from which we can decide if we have well corrected the impulsive noise or not. For this we proceed as follows.

A. Decision Criteria

We choose the optimal threshold δ that minimizes the average risk $\bar{C}$, namely

$$\begin{aligned}\bar{C} &= \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}p_i \int_{\Sigma_i} p(y|H_i)dy \\ &= p_0(C_{01} - C_{00})P_F + p_1(C_{11} - C_{10})P_D \\ &\quad + (C_{00}p_0 + C_{10}p_1)\end{aligned}.$$

C_{ij} is the cost of choosing hypothesis H_j when H_i is true $(i, j = 0, 1)$. Let p_0 and p_1 denote the probability of occurrence of the hypotheses H_0 and H_1, and suppose that we know these *a priori* probabilities. Therefore, the Bayes criteria [34] defines the region Σ_0 and Σ_1 that minimizes the average risk $\bar{C}$. For each $y = \|\underline{S} - \tilde{\underline{S}}\|^2$, we compute the ratio

$$\Lambda(y) = \frac{p_{Y|H_1}(y|H_1)}{p_{Y|H_0}(y|H_0)}.$$

The region Σ_0 consists of y, for which $\Lambda(y) < \Lambda_0$, and Σ_1 of values for which $\Lambda(x) > \Lambda_0$, where the critical value Λ_0 is given by

$$\Lambda_0 = \frac{p_0(C_{01} - C_{00})}{p_1(C_{10} - C_{11})}.$$

In the following, we assume that $C_{00} = C_{11} = 0$.

Now if P_D and P_F are known, then we can use the information given by the ROC curve. However, the Bayes threshold can also be deduced from the ROC curve

$$\Lambda_0 = \frac{dP_D}{dP_F} \tag{22}$$

where Λ_0 is the slope of ROC curve at the point (P_{F_0}, P_{D_0}) [34]. Once the cost and *a priori* probabilities are known, we deduce from the ROC curve the threshold $\delta^{(\tilde{\nu})}$, such that $P_{F_0} = \int_{\delta^{(\tilde{\nu})}}^{+\infty} p_{y|H_0}(Y|H_0)dy$. If $\|\tilde{\underline{S}} - \underline{S}\|^2 < \delta^{(\tilde{\nu})}$, then we can conclude that we have corrected impulsive noise. So, if p_0 and p_1 are known, then we have to choose the threshold Λ_0. From the ROC curve, we remark that when Λ_0 decreases (that means that the slope decreases), $P_D = p(H_1|H_1)$ increases. Or, in our case, we prefer that P_D increases and P_F decreases.

Note that this procedure can also be applied at the beginning of the decoding algorithm. That means if $\|\underline{S}\|^2$ is less than a certain threshold, then we can conclude that no impulsive noise has taken place in the channel. This has the advantage of avoiding the decoding when there is no impulsive noise. So to calculate this threshold, we use the same technique that we have already explained.

B. Combinatorial Test

Let $T = \kappa\tilde{\nu}$, where κ is an integer such that $\kappa > 1$, and in this part, we vary $\tilde{\nu}$ from 1 to r (see Section III). If $y = \|\tilde{\underline{S}} - \underline{S}\|^2 > \delta^{(\tilde{\nu})}$, then for each value of $\tilde{\nu}$, we proceed as follows. Instead of considering the $\tilde{\nu}$ smallest values of $|\tilde{\Lambda}(x)|$ taken on $W_M^{-[1,\ldots,M]}$ that correspond to the impulsive noise location, we take the T smallest values of $|\tilde{\Lambda}(x)|$, and then we compute $\|\tilde{\underline{S}} - \underline{S}\|^2$ for all possible combinations of $\tilde{\nu}$ elements from T elements, until obtaining $\|\tilde{\underline{S}} - \underline{S}\|^2 \leq \delta^{(\tilde{\nu})}$.

VI. Example

Let $M = 64$ and the length of the GI $= 16$ samples long. Among these N carriers, 12 carriers are null (including the middle null carrier and the zeros padded on both ends). Among the remaining $K = 52$ subcarriers, four are fixed pilots carrying known symbols $P_1, \ldots, P_4$, which are at the position {11 25 38 53}, while the other $N = K - 4 = 48$ subcarriers convey the information.

To correct impulsive noise, we use P_1, P_3, P_4, and two zeros (the middle null carrier and the one at the position 59), hence, $\mathcal{A} = \{11\ 32\ 38\ 53\ 59\}$. Then, we selected the syndrome matrix

$$\underline{\underline{S}} = \begin{pmatrix} S_{11} & S_{32} & S_{38} \\ S_{32} & S_{53} & S_{59} \end{pmatrix} \tag{23}$$

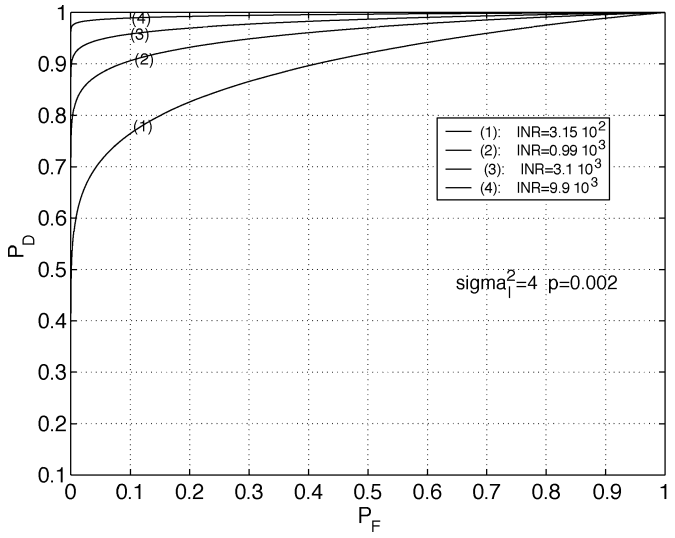


Fig. 5. Theoretical curves ROC.

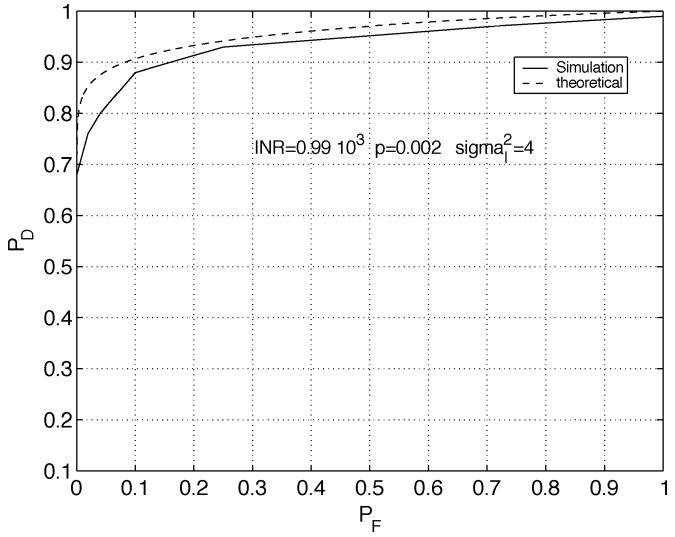


Fig. 6. Comparison between the theoretical ROC curves and simulated ROC curves.

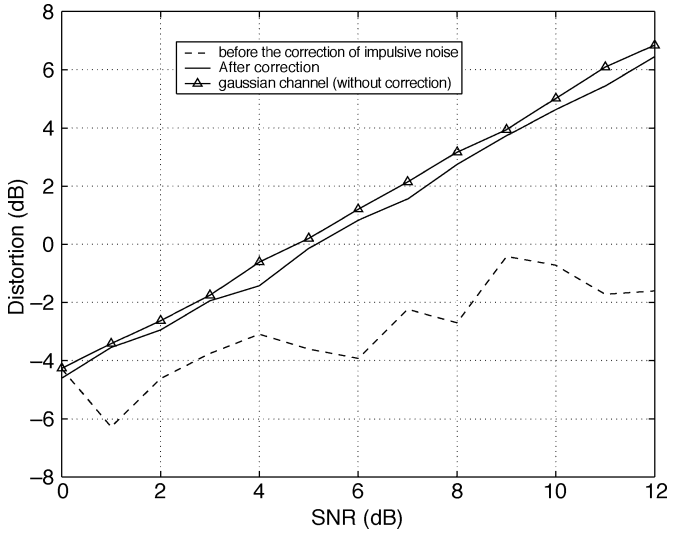


Fig. 7. Distortion performance when we consider a channel, scattered null carriers, and pilot tones.

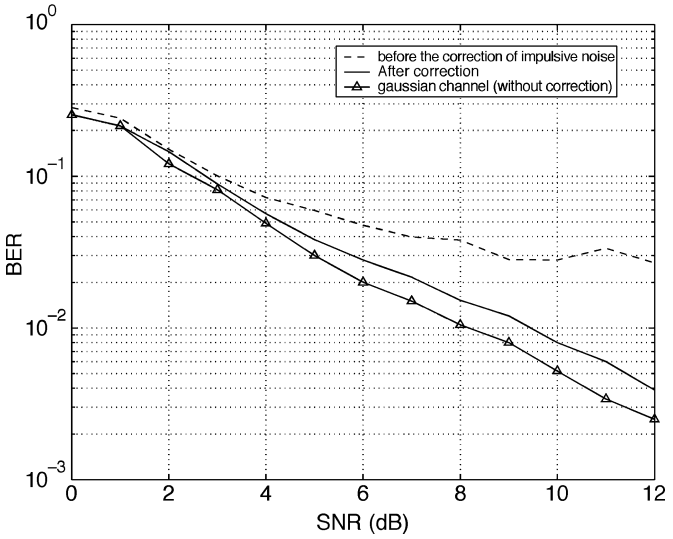


Fig. 8. BER performance when we consider a channel, scattered null carriers, and pilot tones.

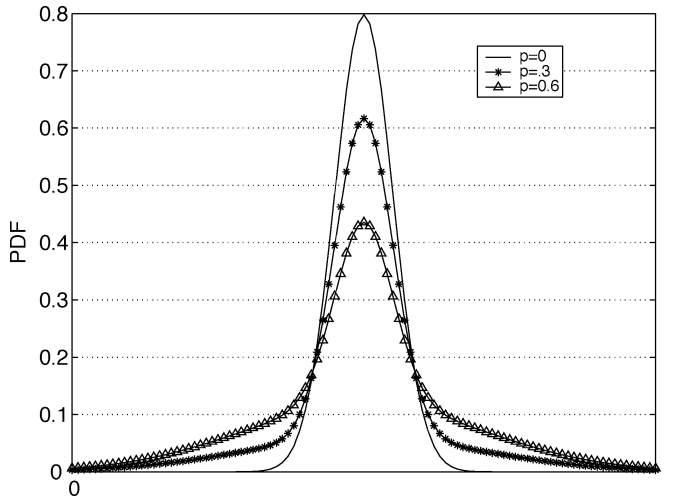


Fig. 9. The pdf of the channel noise.

which can be written as in (10), with $m_0 = 0$, $\delta_0 = 11$, $\theta_0 = 0$, $\theta_1 = 21$, $\delta_1 = 32$, and $\delta_2 = 27$.

It follows that:

$$\underline{\underline{\mathcal{R}}}^{(2)} = \begin{pmatrix} 1 & 1 \\ W_M^{21f_0} & W_M^{21f_1} \end{pmatrix}$$

$$\underline{\underline{\mathcal{Q}}}^{(2)} = \begin{pmatrix} W_M^{11f_0} & W_M^{32f_0} \\ W_M^{11f_1} & W_M^{32f_1} \end{pmatrix}.$$

We verify that $\underline{\underline{\mathcal{R}}}^{(2)}$ and $\underline{\underline{\mathcal{Q}}}^{(2)}$ are invertible because $gcd(\theta_1 - \theta_0, \mathbf{M}) = 1$ and $gcd(\delta_1 - \delta_0, \mathbf{M}) = 1$. Thus we can correct at most two impulsive noises.

We have two hypothesis:

- H_0: there is no impulsive noise;
- H_1: there is impulsive noise.

In the hypothesis H_0, we must consider all these cases:

- I—there is one impulsive noise, and we detect two such that $\underline{\mathrm{loc}} \subset \underline{\tilde{\mathrm{loc}}}$;
- II—there are two impulsive noises, and we detect these;
- III—there are no impulsive noises, and we detect two.

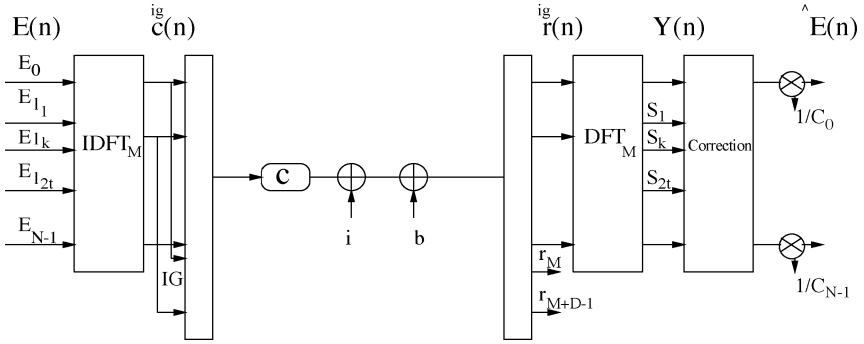


Fig. 10. OFDM system based on the use of a GI.

We have proved in Section V that these three cases are equivalent and have the same pdf $(\|\tilde{\underline{S}} - \underline{S}\|^2/(M\sigma_b^2/2)) \sim \chi^2(2(\beta - \tilde{\nu}))$. Therefore $p_{y|H_0}(y|H_0) = (1/2\sigma_b^6)y\, e^{-(Y/\sigma_b^2)}$ and $P_F = \int_{\Sigma_0} p_{y|H_0}(Y|H_0)dY$.

In the case of the hypothesis H_1, we have to consider all the following possibilities:

- I_1: there is one impulsive noise and we did not detect it;
- I_2: there are two impulsive noises and we detect two, such that only one is at the correct localization;
- I_3: there are two errors and we detect two, such that $\underline{\text{loc}} \bigcap \underline{\tilde{\text{loc}}} = \emptyset$;
- II_k: there are more than two impulsive noises ($2 < k \leq M$), and we can not detect them because the correction capacity is overflowed.

Then

$$P_D = p_{I_1} \int_{\Sigma_1} f(y \in I_1)dy + p_{I_2} \int_{\Sigma_1} f(y \in I_2)dy + p_{I_3} \int_{\Sigma_1} f(y \in I_3)dy + \theta$$

where θ corresponds to the case when the correction capacity is overflowed. The cases I_1, I_2, I_3 and $\{II_k\}_{k>2}$ are with a probability of occurrence denoted $p_{I_1}, p_{I_2}, p_{I_3}$, and p_{II_k}, where

$$p_{I_1} = p_{I_3} = \frac{C_M^1 p^1 (1-p)^{M-1}}{1-(1-p)^M}$$
$$p_{I_2} = \frac{C_M^2 p^2 (1-p)^{M-2}}{1-(1-p)^M}$$
$$p_{II_k} = \frac{C_M^k p^k (1-p)^{M-k}}{1-(1-p)^M} \quad k \in \{3, 4 \ldots M\}.$$

In the following, we neglect the case of overflow. In Fig. 5, we plot the theoretical ROC curves, when $M = 64, \sigma_i = 2$, and $p = 2\ 10^{-3}$ for three different values of INR, where INR $= \sigma_i^2/\sigma_b^2$. The probability p is low, so we neglect the occurrence of three errors. In Fig. 6, we compare the simulated and theoretical ROC curves. We remark that theoretical curves and those obtained by simulation are very close, and this is due to the estimates made to calculate the pdf.

VII. Simulations

These simulations are reminiscent of the Hiperlan2 standard when 4-QAM symbols are transmitted. Low-level Gaussian noise samples with variance σ_b^2 are added to each position independently modeling the background noise. The parameter of the Bernoulli sequence is $p = 10^{-3}$, and the variance of the impulsive noise is $\sigma_i = 70 * \sigma_b$. We have chosen $C_{01}/C_{10} = 1/8$.

Remember that in Hiperlan2, the number of carriers is $M = 64$ and the length of the GI is 16. Among these carriers, 12 are null (including the middle null carrier and the zeros padded on both ends). Among the remaining $K = 52$ subcarriers, four are fixed pilots carrying known symbols P_1, P_2, P_3, and P_4, which are at positions {11 25 39 53} while the other carriers convey the information. To correct impulsive noise, we have used P_1, P_3, P_4, and two zeros (the zeros that is in the middle (32) and one zero on the side band (59)), and where we propose to change the position of P_4 to 38 in order to have condition (17) verified.

In Fig. 7, we plot $1/$MSE (dB) (where MSE is the mean square error) as a function of E_s/N_0 (dB), before and after decoding. We calculate the MSE between the transmitted and the received symbols for four cases: 1) after impulsive noise correction; 2) after adding *a posteriori* control; 3) before impulsive noise correction; and 4) we consider only Gaussian noise. We notice a clear improvement of the performances after using *a posteriori* control. Comparing the case after correction and improvement with the case after correction, one sees that we have a gain of almost 2 dB. However, it is interesting to use *a posteriori* control.

Fig. 8 shows the performances in terms of BER, when we included a channel C which is a realization of the typical channel Model A specified by Hiperlan2. Fig. 9 shows the pdf of the channel noise, and Fig. 10 gives an OFDM system based on the use of a GI. The *a posteriori* control algorithm also shows good behavior under these circumstances, since the curve after correction of the impulsive noise is only marginally different from the curve obtained with Gaussian noise only. The improvement in terms of BER is also shown. Note that this simulation did not contain any classical channel coder. Note also that due to the different situation (zeros are not consecutive), a comparison with the result of [10]–[13] would be very difficult.

VIII. Conclusions

In this paper, we have generalized the procedure of impulsive noise correction to the case when syndromes are scattered among the transmitted sequence. Pilot tones are generally transmitted for synchronization or channel estimation, and can also be seen as additional syndromes and used to correct impulse noise. However, the correction capacity is conditioned by the position of these pilot tones in the transmitted sequence. We have explained the case when capacity is two and the case when it is three.

Classically, the impulsive noise correction is in three steps: 1) estimate the number of impulsive noise; 2) find the impulsive noise locations; and 3) correct the errors. In this paper, we have described the *a posteriori* control step that we have added in order to carefully detect the malfunctions of the decoding algorithm. This presented procedure is essentially based on the theory of hypotheses test. Many extensions are under consideration, in order to increase the practical usefulness of this technique.

References

[1] J. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, no. 28, pp. 14–15, Apr. 1990.

[2] M. Alard and R. Lassalle, "Principles of modulation and channel coding for digital broadcasting for mobile receivers," *EBU Rev.*, pp. 168–190, Aug. 1997.

[3] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665–675, Jul. 1985.

[4] P. S. Chow, J. C. Tu, and J. M. Cioffi, "Performance evaluation of a multichannel transceiver system for ADSL and HDSL services," *IEEE J. Sel. Areas Commun.*, vol. 9, pp. 909–919, Aug. 1991.

[5] K. J. Kerpez, "Forward error correction for asymmetric digital subscriber lines (ADSL)," in *Proc. IEEE Global Telecommun. Conf.*, vol. 3, Phoenix, AZ, 1991, pp. 1974–1978.

[6] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 11, pp. 991–1001, Sep. 1993.

[7] J. G. Proakis, *Digital Communication*. New York: McGraw-Hill, 1989.

[8] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Trans. Commun.*, vol. 44, pp. 145–147, Feb. 1996.

[9] F. Abdelkefi, A. Gabay, and P. Duhamel, "Impulse noise cancellation in multicarrier transmission," in *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, vol. 4, May 2001, pp. 2381–2384.

[10] J. K. Wolf, "Redundancy, the discrete Fourier transform, and impulse noise cancellation," *IEEE Trans. Commun.*, vol. COM-31, pp. 458–461, Mar. 1983.

[11] G. R. Redinbo, "Decoding real block codes: Activity detection, Wiener estimation," *IEEE Trans. Inf. Theory*, vol. 46, pp. 609–623, Mar. 2000.

[12] J.-L. Wu and J. Shiu, "Discrete cosine transform in error control coding," *IEEE Trans. Commun.*, vol. 43, pp. 1857–1861, May 1995.

[13] R. Kumaresan, "Rank reduction techniques and burst error-correction decoding in real/complex fields," in *Proc. Asilomar Conf, Circuits, Syst., Comput.*, Nov. 1985.

[14] F. Marvasti, M. Hasan, M. Echhart, and S. Talebi, "Effecient algorithms for burst error recovery using FFT and other transform kernels," *IEEE Trans. Signal Process.*, vol. 47, pp. 1065–1075, Apr. 1999.

[15] J.-L. Wu and J. Shiu, "Real-valued error control coding by using DCT," *IEE Proc.—I*, vol. 139, no. 2, Apr. 1992.

[16] G. R. Redinbo, "A fault-tolerant decoding procedure for real cyclic codes," in *Proc. Pacific Rim Int. Symp. Fault-Tolerant Syst.*, Taipei, Taiwan, Dec. 1997, pp. 35–40.

[17] C. R. P. Hartmann, "Generalizations of the BCH bound," *Inf., Control*, no. 20, 1972.

[18] C. Roos, "A generalization of the BCH bound for cyclic codes, including the Hartmann–Tzeng bound," *J. Combin. Theory*, ser. A 33, 1982.

[19] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 460–473, Apr. 1972.

[20] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 14–20, Jan. 1972.

[21] F. D. Neeser and J. L. Massey, "Proper complex random processes with application to information theory," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1293–1303, Jul. 1993.

[22] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*. New York: Oxford Univ. Press, 1992.

[23] R. E. Blahut, "Transform techniques for error control codes," *IBM J. Res. Develop.*, vol. 23, pp. 299–315, May 1979.

[24] R. E. Blahut, *Algebraic Methods for Signal Processing and Communications Coding, Signal Processing and Digital Filtering*, C. S. Burrus, Ed. New York: Spring-Verlag, 1992.

[25] ——, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1987.

[26] T. G. Marshall, "Decoding of real-number error-correction codes," in *Proc. GLOBECOM*, San Diego, CA, Nov. 1983, pp. 1249–1303.

[27] C. R. P. Hartmann, K. K. Tzeng, and R. T. Chien, "Some results on the minimum distance structure of cyclic codes," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 402–409, May 1972.

[28] F. Abdelkefi, P. Duhamel, and F. Alberge, "On the use of pilot tones for impulse noise cancellation in Hiperlan2," in *Proc. Int. Symp. Signal Process. Applicat.*, Aug. 2001, pp. 591–594.

[29] ——, "Codage correcteur d'erreur dans le corps des complexes et systèmes multiporteuses," in *Proc. GRETSI*, 2001.

[30] ——, "*A posteriori* control of complex Reed–Solomon decoding with application to impulse noise cancellation in Hiperlan2," in *Proc. Int. Conf. Commun.*, vol. 2, Apr. 2002, pp. 659–663.

[31] O. Rioul, "A spectral algorithm for removing salt and pepper from images," in *Proc. IEEE Dig. Signal Process. Workshop*, Loen, Norway, Sep. 1996, pp. 275–278.

[32] A. Gabay, "Spectral interpolation coder for impulse noise cancellation over a binary symmetric channel," in *Proc. EUSIPCO*, 2000.

[33] S. M. Kay and A. K. Shaw, "Frequency estimation by principal component AR spectral estimation method without eigendecomposition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 95–101, Jan. 1988.

[34] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Spring-Verlag, 1994.

Fatma Abdelkefi received the Engineer Degree in electrical engineering from the Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunis, Tunisia, in 1998, the M.S. degree in digital communications systems in 1999, and the Ph.D. degree with highest honors in 2002, both from the Ecole Nationale Supéieure de Télécommunications (ENST), Paris, France.

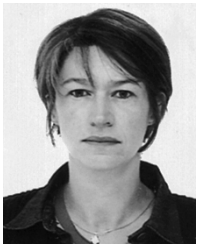
She joined the Signals and Systems Laboratory of SUPELEC, Gif-sur-Yvette, France, as a Research Assistant. In November 2002, she joined the Signal Processing Group of the Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA), Paris, France, as a Research Associate. Since November 2004, she has been with the Communication Technology Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland. Her research interests include digital communications, information theory, detection and estimation, equalization, and multicarrier systems.

Dr. Abdelkefi was awarded a Government Merit Scholarship to conduct her doctoral studies in France in the field of electrical engineering (signal processing and digital communications).

Pierre Duhamel (F'98) was born in France in 1953. He received the Ing. degree in electrical engineering from the National Institute for Applied Sciences (INSA) Rennes, France, in 1975, and the Dr.Ing. degree and the Doctoratèes sciences degree in 1978 and 1986, respectively, both from Orsay University, Orsay, France.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive CCD filters. Later, he worked on fast algorithms for computing Fourier transforms and convolutions, and applied similar techniques to adaptive filtering, spectral analysis and wavelet transforms. From 1993 to September 2000, he was a Professor at the National School of Engineering in Telecommunications (ENST), Paris, France, with research activities focused on signal processing for communications. He was Head of the Signal and Image Processing Department from 1997 to 2000. He is now with CNRS/LSS (Laboratoire de Signaux et Systemes), Gif-sur-Yvette, France, where he is developing studies in signal processing for communications (including equalization, iterative decoding, multicarrier systems) and signal/image processing for multimedia applications, including source coding, joint source/channel coding, watermarking, and audio processing.

Dr. Duhamel was Chairman of the Digital Signal Processing Committee from 1996 to 1998, was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, and was Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He was a Guest Editor for the special issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING on wavelets. He was an IEEE Distinguished Lecturer for 1999, and was Co-General Chair of the 2001 International Workshop on Multimedia Signal Processing, Cannes, France. The paper on subspace-based methods for blind equalization, which he coauthored, received the Best Paper Award from the IEEE Signal Processing Society in 1998. He was awarded the "Grand Prix France Telecom" award by the French Science Academy in 2000.



Florence Alberge was born in Albi, France, in 1971. She received the Ingenieur Degree in electrical engineering from the Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA), Cergy-Pontoise, France, in 1996, and the Ph.D. degree from the Ecole Nationale Supéieure de Télécommunications (ENST), Paris, France, in 1999.

Since September 2000, she has been an Assistant Professor with Orsay University, Orsay, France. Her research interests are in the area of signal processing for digital communications, adaptive filtering, and wireless networks.

Annexe 2

F. Alberge, M. Nikolova, P. Duhamel, Blind Identification/Equalization Using Deterministic Maximum Likelihood and a Partial Prior on the Input. IEEE Trans. on Signal Processing, vol 54, n° 2, 2006.

Abstract

A (semi)deterministic maximum likelihood (DML) approach is presented to solve the joint blind channel identification and blind symbol estimation problem for single-input multiple-output systems. A partial prior on the symbols is incorporated into the criterion which improves the estimation accuracy and brings robustness toward poor channel diversity conditions. At the same time, this method introduces fewer local minima than the use of a full prior (statistical) ML. In the absence of noise, the proposed batch algorithm estimates perfectly the channel and symbols with a finite number of samples. Based on these considerations, an adaptive implementation of this algorithm is proposed. It presents some desirable properties including low complexity, robustness to channel overestimation, and high convergence rate.

Blind Identification/Equalization Using Deterministic Maximum Likelihood and a Partial Prior on the Input

Florence Alberge, Mila Nikolova, and Pierre Duhamel, *Fellow, IEEE*

***Abstract*—A (semi)deterministic maximum likelihood (DML) approach is presented to solve the joint blind channel identification and blind symbol estimation problem for single-input multiple-output systems. A partial prior on the symbols is incorporated into the criterion which improves the estimation accuracy and brings robustness toward poor channel diversity conditions. At the same time, this method introduces fewer local minima than the use of a full prior (statistical) ML. In the absence of noise, the proposed batch algorithm estimates perfectly the channel and symbols with a finite number of samples.**

Based on these considerations, an adaptive implementation of this algorithm is proposed. It presents some desirable properties including low complexity, robustness to channel overestimation, and high convergence rate.

***Index Terms*—Adaptive algorithm, blind equalization, deterministic maximum likelihood method, joint estimation, prior knowledge.**

I. Introduction

BLIND identification is an important problem in many areas and especially in wireless communications. Blind techniques present some advantages compared to the traditional training methods [1], [2]. First, the reduced need for overhead information increases the bandwidth efficiency. Furthermore, in certain communication systems, the synchronization between the receiver and the transmitter is not possible; thus training sequences are not exploitable. Finally, even if some training sequence exists, the combination of trained and blind techniques can often lead to improved performances, allowing fast tracking of time-varying channels, for example.

Early approaches to blind equalization were based on higher order statistics of the received signal [3] since the second-order statistics of a scalar system output do not contain enough information to identify a nonminimum phase system. Although these algorithms are robust and reliable in many cases, estimating high-order statistics usually requires a large number of data samples. Hence, their application in fast varying environment is intrinsically limited. Tong *et al.* suggested a different option [4]. They proposed to introduce time or spatial diversity at the output. Then, the system considered is a single-input multiple-output (SIMO) system. The SIMO equalization problem can be solved using second-order statistics only, as long as the subchannels do not share common zeros. In a fast fading environment, the statistical model of the input may not be available, or there may not be enough samples to find a reliable estimate of the statistics. In this kind of scenario, the problem may be solved by treating the input as a deterministic variable. Generally, the resulting methods have the finite sample convergence property (i.e., the channel can be perfectly estimated using a finite number of samples in noiseless situations). This is a desirable property especially in packet transmission systems.

In this paper, we focus on deterministic maximum likelihood (DML) methods since they have the additional advantage of being high signal-to-noise ratio (SNR) efficient [5]. Among the major contributions to DML methods, we can cite the two-step maximum likelihood (TSML) [6] and the iterative quadratic maximum likelihood (IQML) [7], both concentrating on channel estimation. Feder *et al.* proposed in [8] a dual algorithm to IQML which aims at estimating the symbols at each step. Unfortunately, the adaptive implementation of these methods is often cumbersome. Another DML method, the maximum likelihood block algorithm (MLBA), has been proposed in [9]. The MLBA performs least squares estimation both in the channels and in the symbols in an alternating manner. This formulation permits to derive easily an adaptive algorithm (MLAA) as shown by the authors in [10]. The MLAA presents some nice properties including low-complexity in computation. However, it is not robust to the overestimation of the channel order and it has a limited ability to track time-varying channels.

In this paper, we present a new algorithm that meets the following four characteristics: adaptivity, low complexity, good speed of convergence, and robustness to the overestimation of the channel order. The proposed method consists of incorporating prior information (related to the input signal) into the DML criterion. The two first properties follow from the MLAA-like structure of the algorithm and the last characteristics are a consequence of the use of the prior. Seshadri [11] and Gosh and Weber [12] first proposed to incorporate the finite alphabet properties into DML to improve the accuracy of the estimates. Later, Talwar proposed the iterative least square with projection (ISLP) [13], which estimates the symbols first without taking the finite alphabet property into account and then projects the estimates onto the alphabet. The problem with these methods is that their convergence is not guaranteed in general and that the incorporation of the finite alphabet property often increases the number of local minima. This is partially solved here by considering only a partial prior on the symbols in order to limit the number of additional spurious local minima (different from the global one). In the proposed approach, a continuous probability distribution function is used which reflects our prior knowledge on the input.

Manuscript received April 5, 2004; revised March 26, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Helmut Boelcskei.

F. Alberge and P. Duhamel are with Supelec/LSS, 91192 Gif-Sur-Yvette Cedex, France (e-mail: alberge@lss.supelec.fr; pierre.duhamel@lss.supelec.fr).

M. Nikolova is with CMLA-ENS de Cachan and CNRS UMR 8536, 94235 Cachan Cedex, France (e-mail: nikolova@cmla.ens-cachan.fr).

Digital Object Identifier 10.1109/TSP.2005.861787

Using Bayesian theory and the classical likelihood function, we are able to derive a new criterion. Such a criterion has also been used in [14] for the purpose of binary images reconstruction. The proposed relaxation technique can also be found in [15] and [16] for application to multiuser detection in CDMA systems. However, channel estimation is not involved in these two contributions. The new algorithm, called conditional maximum likelihood batch algorithm (CMLBA), is then obtained in the same way as the MLBA (alternating minimization) and the corresponding adaptive version is derived.

Concerning the local minima problem, we prove that for a "weak" prior (to be defined later), the stationary points of the CMLBA are also stationary points of the MLBA. And, for each stationary point $(\hat{\mathbf{h}}, \hat{\mathbf{s}})$ of the MLBA, there exists a scale factor α such that $(\alpha\hat{\mathbf{h}}, \hat{\mathbf{s}}/\alpha)$ is also a stationary point of the CMLBA. Thus, the use of such a prior does not increase the number of local minima and, at the same time, it brings robustness to poor channel diversity conditions, as shown in the experimental results. For a stronger prior, the number of local minima is likely to increase. However, we show below that a local minimum is not stable through a recursive procedure, as already shown in [10]. As a result, the proposed recursive algorithm is unlikely to converge toward a local minimum.

This paper is organized as follows. Section II presents the general setup and some properties about the DML criterion. For noise-free data, we recall that the global minimum of the DML criterion is unique. The derivation of the CMLBA is available in Section III. The local minima problem is analyzed in Section IV. In Section V, we explain how to improve the quality of the estimators in the particular case of an ill-conditioned channel matrix. An adaptive version of the CMLBA is proposed in Section VI. The performance of the algorithms and comparison with existing approaches are provided in Section VII.

II. Problem Formulation

This paper addresses SIMO systems (see Fig. 1). Let $\{\tilde{s}(n)\}$ denote the symbol sequence at the input of the system and $x_i(n), 1 \leq i \leq L$, the ith output. The output $x_i(n)$ may be the signal picked on the ith sensor of an array (spatial diversity); or may be obtained by oversampling of a factor L the continuous time signal received on a single sensor (time diversity); or may follow from the combination of both spatial and time diversity. Such a system is described as

$$\mathrm{x}_i(n) = \sum_{k=0}^{M} \tilde{\mathrm{h}}_i(n-k)\tilde{\mathrm{s}}(k) + \mathrm{b}_i(n)\; i = 1, \ldots, L$$

where $\mathbf{h}_i = [\mathbf{h}_i(0), \ldots, \mathbf{h}_i(M)]^T$ is the channel impulse response, M is the maximum order of any channel, and $\{b_i(n)\}, 1 \leq i \leq L$, is a Gaussian independent identically distributed (i.i.d.) additive noise. Sequences $\{b_i(n)\}$ and $\{b_j(n)\} (i \neq j)$ are assumed uncorrelated. For convenience, we adopt the following notations throughout this paper.

- h, s are variables denoting any channel and any symbol sequence, respectively.

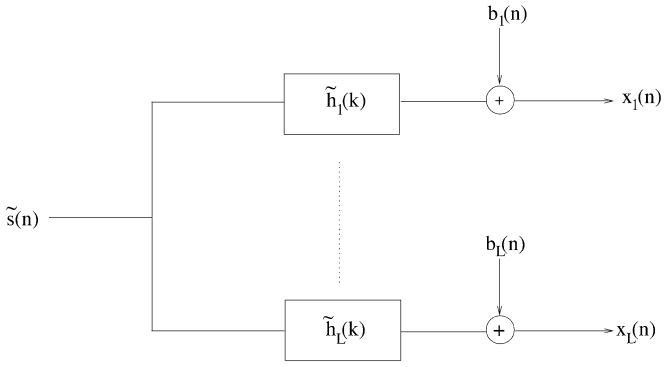


Fig. 1. Single-input/multiple-output system.

- $\tilde{\mathrm{h}}, \tilde{\mathrm{s}}$ are the true channels and the true symbols, respectively, and $\mathbf{x}(n)$ stands for the corresponding observation at time n.
- $\hat{\mathrm{h}}, \hat{\mathrm{s}}$ are the estimates of $\tilde{\mathrm{h}}$ and $\tilde{\mathrm{s}}$.
- $\mathbf{s}_N(n) = [s(n), s(n-1), \ldots, s(n-N-M+1)]^T$ and n is the time index.
- $\hat{\mathbf{s}}_N^{(i)}(n+i) = [\hat{s}(n+i), \hat{s}(n+i-1), \ldots, \hat{s}(n+i-N-M+1)]^T$ is a vector of length $M+N$ containing symbols estimated at iteration i.

The channel impulse responses $\tilde{\mathbf{h}}_i, 1 \leq i \leq L (L > 1)$ are assumed to have a finite length and M stands for the maximum order of any channel. Let $\mathbf{X}_N(n) = [\mathrm{x}_1(n), \ldots, \mathrm{x}_L(n), \ldots, \mathrm{x}_1(n-N+1), \ldots, \mathrm{x}_L(n-N+1)]^T$ denote the vector obtained by interleaving the outputs of the different channels and $\tilde{\mathbf{h}}(k) = [\tilde{h}_1(k), \ldots, \tilde{h}_L(k)]^T$. Then, the output $\mathbf{X}_N(n)$ reads

$$\mathbf{X}_N(n) = \mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N(n) + \boldsymbol{B}_N(n) \tag{1}$$

where $\mathbf{B}_N(n)$ stands for the noise vector. In (1), operator $\mathcal{T}_N$ transforms the set of channel coefficients $\mathbf{h}(k) = [h_1(k), \ldots, h_L(k)]^T, k = 0, \ldots, M$ into the following $LN \times (M+N)$ generalized Sylvester matrix:

$$\mathcal{T}_N(\mathbf{h}) = \begin{pmatrix} \mathbf{h}(0) & \ldots & \mathbf{h}(M) & 0 & \ldots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \ldots & 0 & \mathbf{h}(0) & \ldots & \mathbf{h}(M) \end{pmatrix}.$$

Let $\mathcal{U}$ be the operator that transforms a vector $\mathbf{s}_N(n)$ into an $LN \times L(M+1)$ matrix $\mathcal{U}(\mathbf{s}_N(n))$, in such a way that

$$\mathcal{U}(\mathbf{s}_N(n))\mathbf{h} = \mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n), \quad \forall \mathbf{s}_N, \quad \forall \mathbf{h}. \tag{2}$$

It can be shown that this matrix reads

$$\mathcal{U}(\mathbf{s}_N(n)) = \begin{pmatrix} I_L \bigotimes s_1(n)^T \\ I_L \bigotimes s_1(n-1)^T \\ \vdots \\ I_L \bigotimes s_1(n-N+1)^T \end{pmatrix}$$

where $\bigotimes$ stands for the Kronecker product and I_L is the $L \times L$ identity matrix. The results displayed in the paper rely on the following assumptions.

H1) $\mathcal{T}_N(\tilde{\mathbf{h}})$ is full column rank.

H2) The symbol sequence $\tilde{s}_N(n)$ has linear complexity $2M+1$ or greater [17]. The linear complexity of the sequence $\{\tilde{s}(n-k)\}_{k=0}^{k=N+M-1}$ is defined as the smallest value of c for which there exists $\{\lambda_j\}_{j=1}^{c}$ such as

$$\tilde{s}(n-i) = -\sum_{j=1}^{c} \lambda_j \tilde{s}(n-i-j) \quad i = c, \ldots, N+M-1.$$

The linear complexity measures the predictability of a finite length deterministic sequence.

H2') When H2) is met, it can be shown that $\mathcal{U}(\tilde{\mathbf{s}}_N(n))$ is full column rank.

H3) M (maximum order of the channels) is known or correctly estimated.

H4) The emitted symbols belong to a phase-shift keying (PSK) modulation.

Assumption H1) ensures that (1) is an overdetermined system for $\tilde{\mathbf{h}}$ fixed. This assumption is most often met. However, situations with poor channel diversity conditions may occur. So, it is important to develop methods that are robust to this situation. Similarly, H2') ensures that (1) is an overdetermined system for $\tilde{\mathbf{s}}_N(n)$ fixed. Denote $\nu_M(\tilde{\mathbf{s}}_N(n))$ as the matrix defined by the equation at the bottom of the page. Then, H2) implies $\text{rank}(\nu_M(\tilde{\mathbf{s}}_N(n))) = 2M+1$. Hence, the sample covariance of the vector sequence $\tilde{\mathbf{s}}_{M+1}(n) = [\tilde{s}(n), \tilde{s}(n-1), \ldots, \tilde{s}(n-2M)]^T$ is full rank [18]. We can remark that $\text{rank}(\nu_M(\tilde{\mathbf{s}}_N(n)) = 2M+1$ implies that, necessarily, $N \geq 3M+1$.

The problem considered in this paper is to identify both $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{s}}_N(n)$ based on $\mathbf{X}_N(n)$ only.

The blind equalization problem is viewed as a joint channel and symbol estimation problem. The criterion used is the DML criterion. Following [6], [9], and [7], $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N)$ are estimated through the minimization of the DML criterion with respect to the joint variable $(\mathbf{h}, \mathbf{s}_N(n))$

$$\begin{aligned} \mathcal{J}(\mathbf{h}, \mathbf{s}_N(n)) &= ||\mathbf{X}_N(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n)||^2 \\ &= ||\mathbf{X}_N(n) - \mathcal{U}(\mathbf{s}_N(n))\mathbf{h}||^2. \end{aligned} \tag{3}$$

Hence, the estimated channel and symbols read

$$(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n)) = \arg \min_{(\mathbf{h}, \mathbf{s}_N(n))} \mathcal{J}(\mathbf{h}, \mathbf{s}_N(n)). \tag{4}$$

In the noiseless case, the global minimum is obtained only for the true channel and symbols (up to a scale factor) [9], [19].

III. Conditional Maximum Likelihood Technique

In a fast fading environment, building reliable statistical estimates is a problem: data related to a given channel are not numerous. In such a situation, the symbols are assumed arbitrary and a deterministic method is used. But, if the system is not time-varying and if the data sequence is long enough so that the statistical estimates are reliable, then a statistical method should be used, since in that case the statistical method outperforms the deterministic one in terms of estimation accuracy. The approach proposed in this section is a tradeoff between DML and SML with the additional advantage that it can be used either when the channel is time-varying or not. In this approach, we consider the transmitted symbols to be no longer deterministic quantities but random variables that obey to an arbitrary (different from the true) statistical distribution. As a result, the obtained algorithm involves a lower computational cost than the statistical method and provides a better estimation accuracy than a DML method.

A. Derivation of the Constrained Criterion

A full use of the knowledge on the emitted symbols (their alphabet) usually introduces many local minima. Instead, we propose to consider only a partial prior. Assume that the emitted symbols belong to a PSK modulation, and consider the probability density function (pdf) shown at the bottom of the page, where Z is a normalization term, κ is a positive scalar, and $s(k)$ is the kth component of $\mathbf{s}$. The shape of the distribution function corresponding to $\kappa = 1$ and $\kappa = 10$ is plotted in Fig. 2, where real data are considered. When $s(k)$ is outside the unit circle, the probability is zero, whereas for symbols inside the unit circle, the probability increases with $||s(k)||$. When κ tends to zero, the shape $p(s(k))$ tends to be uniform within the unit circle and zero outside. Even if this special case corresponds to a very "weak" prior, it is of interest, since some properties of the convergence points of the algorithm can be demonstrated under this assumption. This will help in understanding the local minima problem.

Let $p(\mathbf{X}_N(n)|\mathbf{h}, \mathbf{s}_N)$ denote the likelihood function conditioned on both the channels and the symbols. The conditional likelihood function $p(\mathbf{X}_N(n), \mathbf{s}_N|\mathbf{h})$ reads

$$p(\mathbf{X}_N(n), \mathbf{s}_N \,|\, \mathbf{h}) = p(\mathbf{X}_N(n) \,|\, \mathbf{h}, \mathbf{s}_N) p(\mathbf{s}_N). \tag{6}$$

$$\nu_M(\tilde{\mathbf{s}}_N(n) = \begin{pmatrix} \tilde{s}(n) & \tilde{s}(n-1) & \cdots & \tilde{s}(n-N+M+1) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{s}(n-2M) & \tilde{s}(n-2M-1) & \cdots & \tilde{s}(n-N-M+1) \end{pmatrix}$$

$$\begin{cases} p(s(k)) = 0 & \text{if } ||s(k)|| > 1, \\ p(s(k)) = \frac{1}{Z} e^{\kappa ||s(k)||^2}, & \text{if } ||s(k)|| \leq 1 \quad p(\mathbf{s}) = \Pi_k p(s(k)) \end{cases} \tag{5}$$

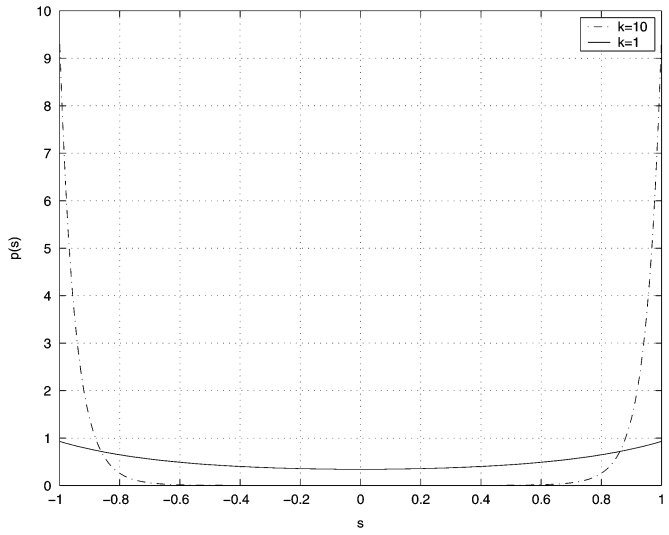


Fig. 2. Probability distribution function for $\kappa = 1, \kappa = 10$. The symbols are assumed to have real values.

$\mathbf{B}_N(n)$ is assumed Gaussian, hence $p(\mathbf{X}_N(n)|\mathbf{h}, \mathbf{s}_N)$ satisfies the following relation:

$$p(\mathbf{X}_N \,|\, \mathbf{h}, \mathbf{s}_N) = \frac{1}{(2\pi\sigma_b^2)^{LN/2}} \times \exp\left(-\frac{1}{2\sigma_b^2}\|\mathbf{X}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2\right). \quad (7)$$

Inserting (5) and (7) into (6), we obtain the conditional likelihood function shown at the bottom of the page, where $\gamma = 2\sigma_b^2\kappa$ and $\mathcal{E}^{M+N}$ stands for the unit disk ($\mathcal{E}^{M+N} = \{\mathbf{s}_N \in \mathbb{C}^{M+N} : \|s(k)\| \leq 1, k = 0 \ldots M+N-1\}$). The maximization of $p(\mathbf{X}_N, \mathbf{s}_N|\mathbf{h})$ is equivalent to the minimization of the following criterion on $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$:

$$\mathcal{L}_\gamma(\mathbf{h}, \mathbf{s}_N) = \|\mathbf{X}(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2 - \gamma\|\mathbf{s}_N\|^2 \quad (9)$$

where $\gamma > 0$. Note that a somewhat similar idea has already been exploited by Papadias in [20], where the transmitted symbols are considered as random variables that obey to a zero-mean Gaussian distribution leading to the criterion

$$\mathcal{L}_{\text{Papadias}}(\mathbf{h}, \mathbf{s}_N) = \|\mathbf{X}(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2 + \beta\|\mathbf{s}_N\|^2 \quad (10)$$

with $\beta > 0$. The same class of criteria is exploited in [15] except that the criterion is minimized with respect to the symbols only. The Gaussian assumption leading to (10) is unrealistic: the symbols close to zero have the highest probability. At the opposite, the pdf in (5) reflects the prior knowledge on the input.

The criterion $\mathcal{L}_\gamma(\mathbf{h}, \mathbf{s}_N)$ is a convex criterion with respect to each variable separately as long as $\gamma \leq \lambda_{\min}$, where $\lambda_{\min}$ stands for the smallest eigenvalue of $\mathcal{T}_N(\mathbf{h})^H\mathcal{T}_N(\mathbf{h})$. In the nonconvex case, the quadratic programming problem subject to linear constraints is NP-complete [21]. Moreover, checking only local optimality in constrained nonconvex programming is NP-hard [22]. This means that the computing time to obtain a solution will grow exponentially with the number of variables. Such a computational cost is unaffordable in the context under study. From now on, we shall only consider the case where $\gamma \leq \lambda_{\min}$. The readers interested by the nonconvex quadratic programming problem can refer to [23]–[25].

B. Implementation of the Method

Many solutions can be proposed to solve the constrained optimization problem in (9). Here, we follow the approach proposed by Gesbert [9] for solving the unconstrained criterion in (3). It presents two major advantages: 1) this formulation is well suited for deriving recursive solutions and 2) the prior information can be incorporated easily, which is not the case with IQML-like approaches. In this approach, a least squares estimation is performed successively in the channel and in the symbols, in an alternating manner. At each step of the iterative procedure, the channels and symbols estimates read

$$\hat{\mathbf{h}}^{(k)} = \left[\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)\right]^{-1} \mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H \mathbf{X}_N(n) \quad (11)$$

$$\hat{\mathbf{s}}_N^{(k)} = \arg\min_{\mathbf{s}_N \in \mathcal{E}^{M+N}} \mathcal{L}_\gamma\left(\hat{\mathbf{h}}^{(k)}, \mathbf{s}_N\right) \text{ with } \mathcal{E}^{M+N} = \{\mathbf{s}_N \in \mathbb{C}^{M+N} : \|s(k)\| \leq 1\} \quad (12)$$

where matrix $\mathcal{U}(\hat{\mathbf{s}}_N^{(k)}(n))$ is assumed to be full rank $\forall k$. This class of algorithm will be referred to as CMLBA$_\gamma$. Each step diminishes the value of the criterion and thus the algorithm converges, however possibly toward a spurious local minima. The optimization problem in (12) is solved by a relaxation method as detailed below.

Let $s(j)$ denote the jth component of $\mathbf{s}_N$. The partial criterion $l_\gamma^{(j)}(s(j))$ reads

$$l_\gamma^{(j)}(s(j)) = s^*(j)(A_j - \gamma)s(j) + s^*(j)B_j + B_j^* s(j) - s^*(j)C_j + C_j^* s(j) \quad \text{with } \|s(j)\| \leq 1$$

$$A_j = T_j^{(k)^H} T_j^{(k)}$$

$$B_j = T_j^{(k)^H} \sum_{l \neq j} T_l^{(k)} s_l$$

$$C_j = T_j^{(k)^H} \mathbf{X}_N(n)$$

$$\begin{cases} p(\mathbf{X}_N, \mathbf{s}_N|\mathbf{h}) = \frac{1}{Z\left(2\pi\sigma_b^2\right)^{LN/2}} \exp\left(-\frac{1}{2\sigma_b^2}\{\|\mathbf{X}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2 - \gamma\|\mathbf{s}_N\|^2\}\right), & \text{if } \mathbf{s}_N \in \mathcal{E}^{M+N} \\ p(\mathbf{X}_N, \mathbf{s}_N|\mathbf{h}) = 0, & \text{elsewhere} \end{cases} \quad (8)$$

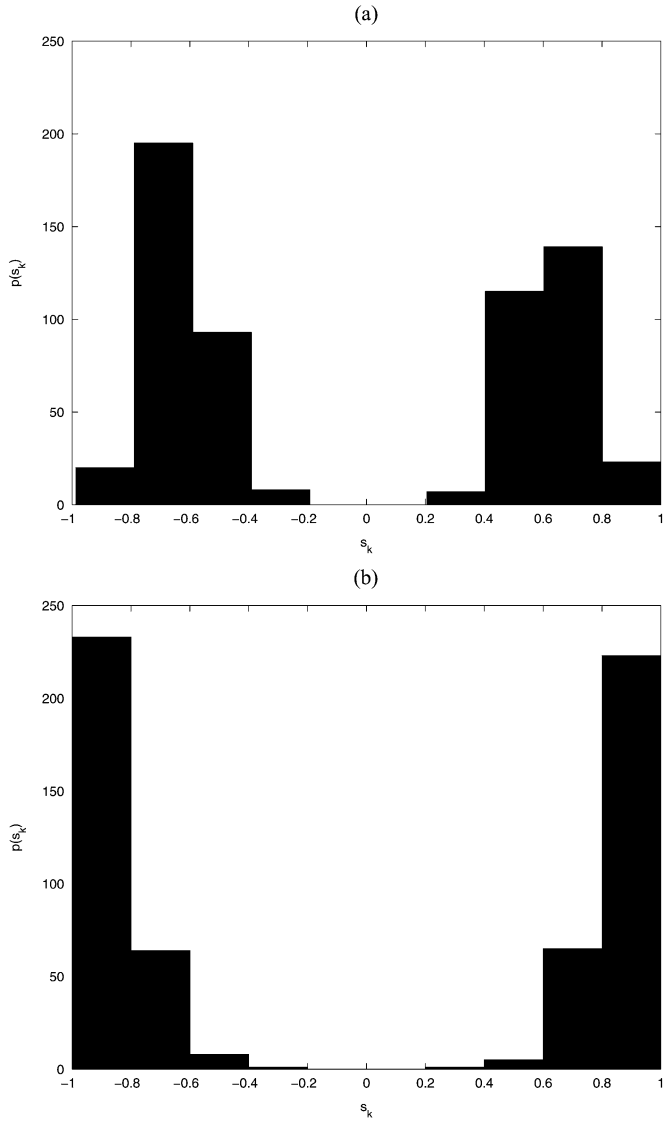


Fig. 3. Histogram of the estimated symbols computed with (a) $\text{CMLBA}_{\min}$ and (b) $\text{CMLBA}_{\max}$. SNR = 10 dB, 600 input symbols.

where $T_j^{(k)}$ is the jth column of $\mathcal{T}_N(\hat{\mathbf{h}}^{(k)})$. The derivative of $l_\gamma^{(j)}(s(j))$ with respect to $s^*(j)$ can be written as

$$\frac{\partial l_\gamma^{(j)}(s(j))}{\partial s^*(j)} = s(j)(A_j - \gamma) + B_j - C_j$$

where $(\partial/\partial s^*(j)) = (1/2)((\partial/\partial t_j) + (\partial/\partial u_j))$ and t_j and u_j stand, respectively, for the real and imaginary part of $s(j)$ [26]. Then at each iteration of the relaxation method, $s(j)$ is computed by the following expression:

$$s(j) = -\mathbf{P}\left(\frac{B_j - C_j}{A_j - \gamma}\right)$$

where $\mathbf{P}$ is the projection operator of $\mathbb{C}^{M+N}$ onto $\mathcal{E}^{M+N}$. The relaxation method converges as long as $(\mathcal{T}_N(\hat{\mathbf{h}}^{(k)})^H\mathcal{T}_N(\hat{\mathbf{h}}^{(k)}) - \gamma\mathbf{I}_{M+N})$ is positive definite [27]. The role played by parameter γ is illustrated in Fig. 3, where we plot the histograms of the symbols estimated via the CMLBA when $\gamma = 0$ (denoted as $\text{CMLBA}_{\min}$) and when $\gamma = \lambda_{\min}$ (denoted as $\text{CMLBA}_{\max}$). A mixed-phase channel is considered and the modulation is a binary PSK (BPSK). It is clear in Fig. 3 that a large value of γ prioritizes the extreme values of the set.

IV. The Local Minima Problem

In this section, the local minima problem is investigated. First, we prove that the global minimum of $\mathcal{L}_\gamma$ is obtained for the true parameters only. Then, a characterization of the local minima is given. The general case $(0 \le \gamma \le \lambda_{\min})$ is first considered, and finally, we concentrate on a special case: the uniform prior on the unit disk $(\gamma = 0)$, which was denoted "weak" prior in the introduction.

A. Uniqueness of the Global Minimum

The following theorem proves the identifiability property for the class of criteria $\mathcal{L}_\gamma(\mathbf{h}, \mathbf{s}_N(n))$.

Theorem 1: In the noiseless case and under H1) and H2), $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ is the global minimum of $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ on $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$ iff $\exists\alpha \in \mathbb{C}^*$ such that $\hat{\mathbf{h}} = \alpha\tilde{\mathbf{h}}, \hat{\mathbf{s}}_N(n) = \tilde{\mathbf{s}}_N(n)/\alpha$. If $\gamma > 0$, then $\|\alpha\| = 1$.

Proof: Stating that $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ is the global minimum of $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ on $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$ is equivalent to the following set of equations:

$$\|\mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n) - \mathbf{X}_N(n)\| = 0 \tag{13}$$

$$\|\hat{\mathbf{s}}_N\|^2 = M + N \ (\text{if}\gamma > 0). \tag{14}$$

In the noiseless case and under H1) and H2), the global minimum is unique up to a scalar factor. Thus, (13) implies that $\exists\alpha \in \mathbb{C}^*$ such that $\hat{\mathbf{h}} = \alpha\tilde{\mathbf{h}}, \hat{\mathbf{s}}_N(n) = \tilde{\mathbf{s}}_N(n)/\alpha$ [9], [19].

The estimated symbols belong to $\mathcal{E}^{M+N}$. Thanks to (14) we conclude that, if $\gamma > 0$, then $\|\alpha\| = 1$. ■

Thus, the global minimum of $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ on the set $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$ is obtained only for the true values of the parameters up to a phase displacement (for $\gamma > 0$) which ensures that the true parameters can be identified.

B. General Case

We first investigate the characterization of the stationary points of the CMLBA_γ when $0 \le \gamma \le \lambda_{\min}$. Let $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ denote a stationary point; then it is solution of the following set of equations:

$$\hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathcal{U}(\hat{\mathbf{s}}_N)]^{-1}\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathbf{X}_N(n) \quad \text{where } \hat{\mathbf{s}}_N \in \mathcal{E}_{M+N} \tag{15}$$

$$\hat{\mathbf{s}}_N(n) = \arg\min_{\mathbf{s}_N \in \mathcal{E}^{M+N}} \mathcal{L}_\gamma(\hat{\mathbf{h}}, \mathbf{s}_N(n)) \quad 0 \le \gamma \le \lambda_{\min}. \tag{16}$$

Equation (16) is a minimization problem subject to inequality constraints. Moreover, $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \mathbf{s}_N(n))$ is a convex function of $\mathbf{s}_N(n)$ as long as $\gamma \le \lambda_{\min}$ and $\mathcal{E}^{M+N}$ is a convex set. Thus, the Kuhn and Tucker [27] relations provide a necessary and sufficient condition for $\hat{\mathbf{s}}_N(n)$ to be a solution of (16). As a result, a stationary point of the CMLBA_γ is such that

$$\hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathcal{U}(\hat{\mathbf{s}}_N)]^{-1}\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathbf{X}_N(n) \quad \text{where } \hat{\mathbf{s}}_N \in \mathcal{E}_{M+N} \tag{17}$$

$$\hat{\mathbf{s}}_N(n) = [\mathcal{T}_N(\hat{\mathbf{h}})^H\mathcal{T}_N(\hat{\mathbf{h}}) - \gamma\mathbf{I}_{M+N} + \Phi]^{-1} \mathcal{T}_N(\hat{\mathbf{h}})^H\mathbf{X}_N(n) \quad \text{where } \hat{\mathbf{s}}_N \in \mathcal{E}_{M+N} \tag{18}$$

$$\Phi = \text{diag}(\phi_i)_{0\le i\le M+N-1}, \quad \phi_i \ge 0 \quad \text{and } \phi_i(\|\hat{s}(n-i)\|^2 - 1) = 0 \;\forall\, i = 0,\dots,M+N-1 \quad (19)$$

where $\hat{s}(n-i)$ stands for the ith component of $\hat{\mathbf{s}}_N$. The difference between the stationary points of the MLBA and of the CMLBA$_\gamma$ is due to the term $\Phi - \gamma\mathbf{I}_{M+N}$ appearing in the expression of $\hat{\mathbf{s}}_N(n)$. Then, the value of the taps of Φ is of major importance. They are further characterized in the following proposition.

Proposition 1: Let $(\hat{\mathbf{h}}, \hat{\mathbf{s}}N(n))$ be a stationary point of the CMLBA$_\gamma$ and Φ the matrix defined in (17)–(19). Then the following relation holds:

$$\text{trace}(\Phi) = \gamma\|\hat{\mathbf{s}}_N\|^2. \quad (20)$$

Proof: $(\hat{\mathbf{h}}, \hat{\mathbf{s}}N(n))$ is a stationary point of the CMLBA. Then (17)–(19) imply that the following two relations are met:

$$\hat{\mathbf{h}}^H\mathcal{U}(\hat{\mathbf{s}}_N)^H[\mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n) - \mathbf{X}_N(n)] = 0 \quad (21)$$

$$\hat{\mathbf{s}}_N(n)^H\mathcal{T}_N(\hat{\mathbf{h}})^H[\mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n) - \mathbf{X}_N(n)] = \hat{\mathbf{s}}_N(n)^H(\gamma\mathbf{I}_{M+N} - \Phi)\hat{\mathbf{s}}_N(n). \quad (22)$$

The left terms of the two equations above are strictly equivalent. Thus (21) and (22) boil down to

$$\gamma\|\hat{\mathbf{s}}_N(n)\|^2 = \hat{\mathbf{s}}_N(n)^H\Phi\hat{\mathbf{s}}_N(n) = \sum_{i=0}^{M+N-1}\phi_i\|\hat{s}(n-i)\|^2.$$

Let I be the set defined as $I = \{i = 0,\dots,M+N-1 : \|s(n-i)\|^2 = 1\}$. The conditions in (19) imply that $\phi_i = 0$ for $i \notin I$. Thus

$$\sum_{i\in I}\phi_i = \text{trace}(\Phi) = \gamma\|\hat{\mathbf{s}}_N(n)\|^2.$$

■

This result illustrates the role of the term $\gamma\|\mathbf{s}_N(n)\|$ in the criterion. Actually, for $\gamma > 0$, the relation $\text{trace}(\Phi) = \gamma\|\hat{\mathbf{s}}_N(n)\|^2$ leads to $\text{trace}(\Phi) > 0$ ($\hat{\mathbf{s}}_N(n) = 0$ is not an acceptable solution). Since, by definition $\phi_i \ge 0$, then it exists $i_0 \in [0,\dots M+N-1]$ such that $\phi_{i_0} > 0$ and consequently $\|\hat{s}(n-i_0)\|^2 = 1$. So, when $\gamma > 0$, there is at least one component of the estimated symbol vector that belongs to the unit circle. The parameter γ permits to push the estimated symbols to the frontier of the set. The special case $\gamma = 0$ is considered in the next paragraph.

C. Special Case: Uniform Prior on the Unit Disk

This case is of special interest, since it forces the symbols to belong to the unit disk which brings robustness to poor channel diversity conditions (see Section VII-A). Furthermore, this so-called "weak" prior is shown below not to introduce additional local minima compared to the unconstrained case. When $\gamma = 0$, (20) becomes $\text{trace}(\Phi) = 0$. All the taps of Φ are nonnegative; then we get $\Phi = 0_{M+N}$. Thus, the stationary points $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ of the CMLBA$_{\min}$ (CMLBA$_\gamma$ with $\gamma = 0$) are a solution of the following set of equations:

$$\begin{cases}\hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathcal{U}(\hat{\mathbf{s}}_N)]^{-1}\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathbf{X}_N(n)\;\hat{\mathbf{s}}_N \in \mathcal{E}^{M+N} \\ \hat{\mathbf{s}}_N(n) = [\mathcal{T}_N(\hat{\mathbf{h}})^H\mathcal{T}_N(\hat{\mathbf{h}})]^{-1}\mathcal{T}_N(\hat{\mathbf{h}})^H\mathbf{X}_N(n)\;\hat{\mathbf{s}}_N \in \mathcal{E}^{M+N}.\end{cases} \quad (23)$$

This means that, once the CMLBA$_{\min}$ has converged, the solution defined above belongs to the set of local minima of the MLBA, corresponding to a specific α (scale factor), which reflects our prior knowledge up to some degree. On the other side, the stationary points of the MLBA are a solution of the system of (23) except that $\hat{\mathbf{s}}_N \in \mathbb{C}^{\mathbb{M}+\mathbb{N}}$. If $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N)$ is a stationary point of the MLBA with $\hat{\mathbf{s}}_N \notin \mathcal{E}^{M+N}$, then $(\alpha\hat{\mathbf{h}}, \hat{\mathbf{s}}_N/\alpha)$ with $\alpha = \max(\|\hat{s}(n-i)\|)_{0\le i\le M+N-1}$ is a stationary point of the CMLBA. Hence, the constraint given by the "weak" prior does not add any local minima to the algorithm. The difference in the set of local minima between the MLBA and the CMLBA lies in the value of the scale factor.

V. CMLBA$_\gamma$ and Ill-Conditioned Filtering Matrix

When some roots of the L subchannel impulse responses are close to each other, the corresponding Sylvester matrix $\mathcal{T}_N(\mathbf{h})$ is hardly full column rank. Thus, $\lambda_{\min}$ (smallest eigenvalue of $\mathcal{T}_N(\mathbf{h})^H\mathcal{T}_N(\mathbf{h})$) is almost zero. In that case, CMLBA$_{\max}$ is equivalent to CMLBA$_{\min}$ and $p(s_k)$ is a uniform pdf. In this section, we explain how we can introduce a strong prior information even when the filtering matrix is badly conditioned. The key point of the method is given by Theorem 2.

Theorem 2: Let $\mathbf{A_1}$ and $\mathbf{A_2}$ be two submatrices of a matrix $\mathbf{A}$ such that $\mathbf{A} = [\mathbf{A_1}\vdots\mathbf{A_2}]$. Then

$$\lambda^A_{\min} \le \lambda^{A_1}_{\min} \text{ and } \lambda^A_{\min} \le \lambda^{A_2}_{\min}$$

where $\lambda^A_{\min}$ (respectively, $\lambda^{A_i}_{\min}$) stands for the smallest eigenvalue of $\mathbf{A}^H\mathbf{A}$ (respectively, $\mathbf{A_i}^H\mathbf{A_i}, i = 1,2$).

Proof: Proof is obvious. ■

If we consider, in the minimization problem, a partition of $\mathcal{T}_N(\mathbf{h})$ instead of considering the whole matrix, then the reduced minimization problem (in terms of number of variables) is at least as well conditioned as the initial problem. Thus, the proposed method consists in partitioning the symbol vector to be estimated and then estimating alternately each part while the rest is fixed to the value obtained at the previous iteration. For simplicity's sake, we give the update equations in the case where the symbol vector is split into two parts. Generalization to others partitions is straightforward since the subvectors have equal lengths (except for the last one if the length of $\mathbf{s}_N$ is not proportional to the number of partitions) and consecutively ordered elements. Let $\mathbf{s}_N = [\mathbf{u}^H\mathbf{v}^H]^H$ where the length of $\mathbf{u}$ (respectively, $\mathbf{v}$) is N_1 (respectively, N_2). The matrix $\mathcal{T}_N(\mathbf{h})$ is also split into two submatrices as: $\mathcal{T}_N(\mathbf{h}) = [\mathcal{T}_N^{1\to N_1}(\mathbf{h})\;\mathcal{T}_N^{N_1\to N_1+N_2}(\mathbf{h})]$ ($\mathcal{T}_N^{1\to N_1}(\mathbf{h})$ contains the N_1 first columns of $\mathcal{T}_N(\mathbf{h})$). Then, at each step of the iterative procedure, the channel and symbol estimates can be read

$$\hat{\mathbf{h}}^{(k)} = \left[\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)\right]^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H\mathbf{X}_N(n) \quad (24)$$

$$\hat{\mathbf{u}}^{(k)} = \arg\min_{\mathbf{u}\in\mathcal{E}^{N_1}}\mathcal{L}_{\gamma_1}\left(\hat{\mathbf{h}}^{(k)}, \begin{bmatrix}\mathbf{u}\\ \hat{\mathbf{v}}^{(k-1)}\end{bmatrix}\right) \quad (25)$$

$$\hat{\mathbf{v}}^{(k)} = \arg\min_{\mathbf{v}\in\mathcal{E}^{N_2}}\mathcal{L}_{\gamma_2}\left(\hat{\mathbf{h}}^{(k)}, \begin{bmatrix}\hat{\mathbf{u}}^{(k)}\\ \mathbf{v}\end{bmatrix}\right) \quad (26)$$

$$\hat{\mathbf{s}}_N^{(k)} = \left[\left(\hat{\mathbf{u}}^{(k)}\right)^H\left(\hat{\mathbf{v}}^{(k)}\right)^H\right]^H \quad (27)$$

with $\gamma_1 = \lambda_{\min}((\mathcal{T}_N^{1\to N_1}(\hat{\mathbf{h}}))^H \mathcal{T}_N^{1\to N_1}(\hat{\mathbf{h}}))$ and $\gamma_2 = \lambda_{\min}((\mathcal{T}_N^{N_1\to N_1+N_2}(\hat{\mathbf{h}}))^H \mathcal{T}_N^{N_1\to N_1+N_2}(\hat{\mathbf{h}}))$. The relevance of the method, in this context, is demonstrated in Section VII.

VI. Adaptive Algorithms

In wireless communication, the channel is time-varying. Thus it is important to develop algorithms able to track those variations. This section is devoted to the derivation of an adaptive algorithm based on CMLBA$_\gamma$.

A. Weighted Criterion

The adaptivity property is obtained by introducing an exponential weighting factor into the definition of the criterion $\mathcal{L}_\lambda$. Let $\mathcal{L}_\lambda^{(W)}$ denote the weighted criterion defined as

$$\mathcal{L}_\lambda^{(W)}(\mathbf{h},\mathbf{s}_N) = \sum_{t=n-N+1}^{t=n} \lambda^{n-t}\|\mathbf{X}_1(t) - \mathcal{T}_1(\mathbf{h})\mathbf{s}_1\|^2 - \gamma\|\mathbf{s}_N\|^2 \quad (28)$$

where $\lambda \in [0;1]$ is the forgetting factor, which ensures that old data are forgotten. Using a matrix formulation, we obtain

$$\mathcal{L}_\lambda^{(W)}(\mathbf{h},\mathbf{s}_N) = \left\|\Lambda_N^{1/2}[\mathbf{X}_N(n+k) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N]\right\|^2 - \gamma\|\mathbf{s}_N\|^2 \quad (29)$$

where $\Lambda_N = \mathrm{diag}([\underbrace{1\ldots 1}_{L}\vdots\underbrace{\lambda\ldots\lambda}_{L}\vdots\ldots\vdots\underbrace{\lambda^{N-1}\ldots\lambda^{N-1}}_{L}])$. Note that the forgetting factor is not applied to $\gamma\|\mathbf{s}_N\|^2$ since this term is related to the prior knowledge which is not time-dependent.

B. Derivation of CMLAAγ

The proposed algorithm is such that the updated estimates of channels and symbols at iteration k are calculated based on both their estimates at iteration $k-1$ and newly arrived data. The proposed approach was first presented by the authors in [10], where the recursive and adaptive versions of the MLBA are derived. The outlines of the method are recalled below, and the update equations for the CMLAAγ are then presented.

Let $\hat{\mathbf{h}}^{(k)}$ and $\hat{\mathbf{s}}_{N+k}^{(k)}(n+k)$ denote, respectively, the channel and the $N+M+k\times 1$ symbol vector estimated at iteration k. The adaptive algorithm is obtained from a growing window procedure after the following simplifications.

S1) The iterative minimization with respect to the joint variable is replaced by a minimization with respect to each variable separately. So, at step k, we compute

$$\hat{\mathbf{s}}_{N+k}^{(k)}(n+k) = \arg\min_{\mathbf{s}_{N+k}\in\mathcal{E}^{M+N+k}} \mathcal{L}_\gamma^{(W)}\left(\hat{\mathbf{h}}^{(k-1)},\mathbf{s}_{N+k}\right) \quad (30)$$

$$\hat{\mathbf{h}}^{(k)} = \arg\min_{\mathbf{h}} \mathcal{L}_\gamma^{(W)}\left(\mathbf{h},\hat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right). \quad (31)$$

S2) At iteration k, (30) updates $M+N+k$ symbols. Hence, the computational complexity involved in (31) grows with k. Here, we propose to compute, at iteration k, the new emitted symbol, and we also update the next P (independent of k) symbols in the delay line where P is a crucial parameter to be determined. Implicitly, the previous symbols are supposed correctly estimated, which is often met since no decision device is introduced. Then (30) is replaced by

$$\begin{aligned}\hat{\mathbf{s}}_{P+1-M}^{(k)}(n+k) &= \arg\min_{Z\in\mathcal{E}^{P+1}} \mathcal{L}_\gamma^{(W)} \times \left(\hat{\mathbf{h}}^{(k-1)}, \begin{bmatrix}\mathbf{Z}\\ \hat{\mathbf{s}}_0^{(k-1)}(n+k-P-1)\end{bmatrix}\right) \quad (32)\\ &= \arg\min_{Z\in\mathcal{E}^{P+1}} \left\|\Lambda_{P+1}^{1/2}\left[\mathbf{X}_{P+1}(n+k) - \mathcal{T}_{P+1}(\mathbf{h}) \times \begin{bmatrix}\mathbf{Z}\\ \hat{\mathbf{s}}_0^{(k-1)}(n+k-P-1)\end{bmatrix}\right]\right\|^2 \\ &\quad - \gamma\left\|\begin{bmatrix}\mathbf{Z}\\ \hat{\mathbf{s}}_0^{(k-1)}(n+k-P-1)\end{bmatrix}\right\|^2. \quad (33)\end{aligned}$$

Note that for the minimization problem in (33), the maximum value of γ is the minimum eigenvalue of $(\mathcal{T}_{P+1}^{1\to P+1})^H\mathcal{T}_{P+1}^{1\to P+1}$, where $\mathcal{T}_{P+1}^{1\to P+1}$ is the submatrix of $\mathcal{T}_{P+1}$ containing its P+1 first columns. Remember that, according to Theorem 2, $\lambda_{\min}(\mathcal{T}_{P+1}^{1\to P+1})^H\mathcal{T}_{P+1}^{1\to P+1}) \geq \lambda_{\min}(\mathcal{T}_{P+1}^H\mathcal{T}_{P+1})$. Then, even if $\mathcal{T}_{P+1}$ is badly conditioned, γ is not necessarily close to zero.

S3) The estimated channel $\hat{\mathbf{h}}^{(k)}$ is updated recursively from $\hat{\mathbf{h}}^{(k-1)}$, which is done without any approximation.

In the following, we derive the update equations for CMLAA$_\gamma$ (CML *adaptive* algorithm). We consider separately the minimization with respect to the symbols and the minimization with respect to the channel.

1) Minimization With Respect to the Symbols: The optimization problem in (33) is solved by a relaxation method. Let $s(j)$ denote the jth component of $\hat{\mathbf{s}}_{P+1-M}^{(k),(i)}(n+k)$, where k is the iteration number of CMLAA$_\gamma$ and i is the iteration number of the relaxation method; and let $T_{j,\lambda}^{(k-1)}$ denote the jth column of $\Lambda_{P+1}^{1/2}\mathcal{T}_{P+1}(\hat{\mathbf{h}}^{(k-1)})$. Then $\hat{\mathbf{s}}_{P+1-M}^{(k)}(n+k)$ is the stationary point obtained through the following iterative algorithm:

```
While ||ŝ^{(k),(i)}_{P+1−M}(n+k) − ŝ^{(k),(i−1)}_{P+1−M}(n+k)|| > ε do
  For j = 1 : P + 1
    A_j = T^{(k−1)H}_{j,λ} T^{(k−1)}_{j,λ}
    B_j = T^{(k−1)H}_{j,λ} Σ_{l=1,l≠j}^{l=P+1+M} T^{(k−1)}_{l,λ} s(l)
    C_j = T^{(k−1)H}_{j,λ} Λ^{1/2}_{P+1} X_{P+1}(n)
    s(j) = −P((B_j − C_j/A_j − γ))
  end
end
```

2) Minimization With Respect to the Channel: The update of the filter for CMLAA$_\gamma$ can be performed using (34) shown at the bottom of the next page (for more details, see [10]), where $\mathbf{R}^{(i)} = \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))^H\Lambda_{N+i}\mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))$. Simplifications of these equations are provided in [10], where the recursive least squares (RLS)-like algorithm above is turned into a least mean square (LMS)-like algorithm with almost no loss in performance.

C. Initialization

CMLAA$_\gamma$ needs of course a reliable initialization. A similar problem is encountered in TSML [6] or IQML [28]. Generally,

the problem is solved by the use of an initialization procedure such as the subspace algorithm for example. Here, we propose to initialize the CMLAA$_\gamma$ with the solution given by the corresponding batch algorithm to a minimization problem over a block of size N. In any case, we take $N > 3M + 1$ to ensure that $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N)$ is the only global minimum of the considered criterion [see H2)].

D. Properties of CMLAA$_\gamma$

CMLAA$_\gamma$ exhibits some desirable properties for tracking the channel parameters in practical contexts, such as GSM, for example. This properties are summarized below.

- The introduction of the a priori knowledge into the criterion improves the convergence speed (see Section VII-B) of the algorithm as well as its tracking capabilities.
- The computational cost is moderate. Indeed, the most demanding part is the minimization with respect to the symbols. In CMLAA$_\gamma$, the length of the symbol vector to be estimated is P at each iteration (against $N + M$ in the CMLBA). We will see in the simulation part that a good choice for P is the channel order M. The estimation of the channel is performed by an LMS-like algorithm.
- CMLAA$_\gamma$ appears to be robust to the overestimation of the channel (see Section VII-B). This property is mandatory for using the proposed algorithm in practical applications.

E. Convergence of CMLAA$_\gamma$

In this section, we point out a result concerning the convergence of the CMLAA$_\gamma$ established for $\lambda = 1$ (which means that the adaptivity property is lost). The main result is formalized below.

Theorem 3: In the noiseless case, if CMLAA$_\gamma$ $(\lambda = 1)$ converges, if the assumptions H1) and H2) are met, and if for all k situated after the convergence $\hat{\mathbf{s}}_1^{(k)}(n+k) \neq \mathbf{0}_{M+1}$, then CMLAA$_\gamma$ converges toward the global minimum.

Proof: The proof is identical to that of [10, Theorem 3]. ■

This result is the consequence of the stability of the global minimum (instability of a local minimum) during a recursive procedure proved in [10, Theorem 2]. This result has first been established in the noiseless case. However, if we express the DML criterion as a function of the channel only $(\|\mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n) - \mathbf{X}_N(n)\|^2 = \|\mathcal{P}^\perp \mathbf{X}_N(n)\|^2$ where $\mathcal{P}^\perp$ is the projection matrix on the range of $\mathcal{T}_N(\mathbf{h})$), it can be proved that when N tends to infinity, the noisy DML criterion tends to the noiseless one [29]. Then, the proof is also relevant in the noisy case as far as the number of data $(M + N)$ is large enough. In other words, the theorem states that the global minimum is the only stationary point in a recursive procedure.

VII. Simulation

To gain more insights about the results obtained in the previous sections, we present some numerical evaluations. The performance of the algorithms is measured by the normalized root mean square error in decibels

$$\mathrm{NRMSE}_{dB}(\mathbf{h}) = 20\log_{10}\left(\frac{1}{\|\tilde{\mathbf{h}}\|}\sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r}\left\|\hat{\alpha}^{(i)}\hat{\mathbf{h}}^{(i)} - \tilde{\mathbf{h}}\right\|^2}\right)$$

where $\hat{\mathbf{h}}^{(i)}$ stands for the estimated channel from the ith trial, $\tilde{\mathbf{h}}$ is the true channel, and $\hat{\alpha}^{(i)} = \arg\min_\alpha \|\alpha\hat{\mathbf{h}}^{(i)} - \tilde{\mathbf{h}}\|^2$. N_r denotes the number of Monte Carlo runs. Noise samples are generated from i.i.d. zero-mean Gaussian random sequences with variance σ^2. The symbols belong either to a PSK modulation or to a 8-quadrature amplitude modulation (QAM). Figs. 4–7 concern the batch algorithm, whereas Figs. 8–10 concern the adaptive algorithms.

A. Batch Algorithms

In this section, we present some simulation studies of the MLBA and of CMLBA$_\gamma$ where only the extreme values of γ are considered (i.e., $\gamma = 0, \gamma = \lambda_{\min}(\mathcal{T}_N(\hat{\mathbf{h}})^H\mathcal{T}_N(\hat{\mathbf{h}}))$). These two algorithms will be referred to as CMLBA$_{\min}$ and CMLBA$_{\max}$, respectively. The performance of these algorithms is compared against those of TSML [6], the multistep linear prediction algorithm (MLPA) [30], and the (joint order detection and channel estimation via least squares smoothing (J-LSS) [31]. We also consider the method described in Section V. We call it CMLBA$_{\max}(N_P)$, where N_P stands for the number of subvectors of $\mathbf{s}_N$ updated separately [for example, the algorithm in (25) and (26) is CMLBA$_{\max}$ (2)].

1) Description of the Multipath Channels: In our simulations, we considered the following channels commonly used in the literature.

$$\begin{cases}
\hat{\mathbf{h}}^{(i)} = \hat{\mathbf{h}}^{(i-1)} + \left[\mathbf{R}^{(i)}\right]^{-1}\left\{\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \Lambda_{P+1}\left[\mathbf{X}_{P+1}(n+i) - \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\right]\right. \\
\left. -\lambda\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H \Lambda_P\left[\mathbf{X}_P(n+i-1) - \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)}\right]\right\} \\
\left[\mathbf{R}^{(i)}\right]^{-1} = \mathbf{A} + \mathbf{A}\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H\left[\frac{\Lambda_P}{\lambda}\right. \\
\left. -\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\mathbf{A}\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H\right]^{-1} \times \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\mathbf{A} \\
\mathbf{A} = \left[\lambda\mathbf{R}^{(i-1)}\right]^{-1} - \left[\lambda\mathbf{R}^{(i-1)}\right]^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \times [\Lambda_{P+1} \\
+\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\left[\lambda\mathbf{R}^{(i-1)}\right]^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H\Big]^{-1} \times \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\left[\lambda\mathbf{R}^{(i-1)}\right]^{-1}
\end{cases} \tag{34}$$

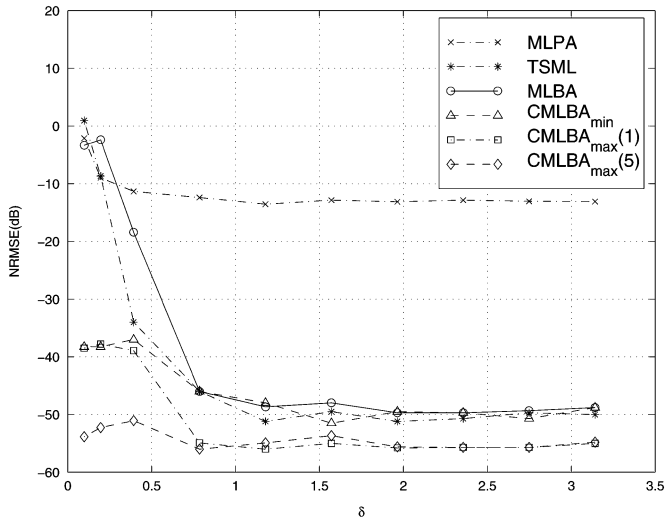


Fig. 4. Performance comparison versus δ (100 Monte Carlo runs, $M+N=32$ BPSK input symbols).

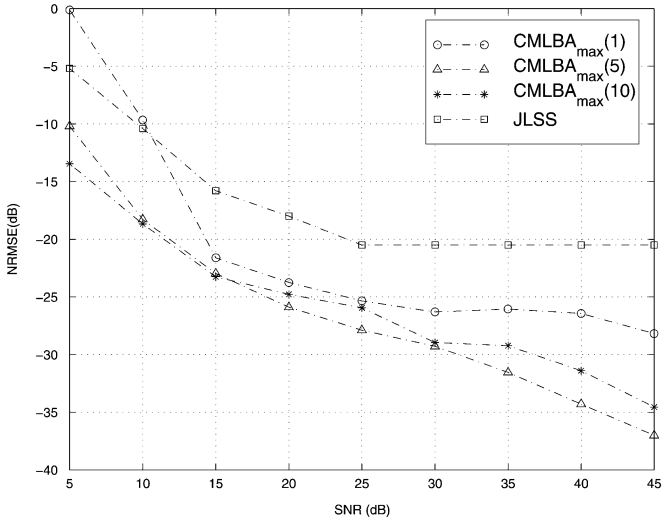


Fig. 6. Performance comparison. $\mathbf{h}^{\mathrm{Tong}}$, 100 Monte Carlo runs, $M+N=100$ QPSK input symbols.

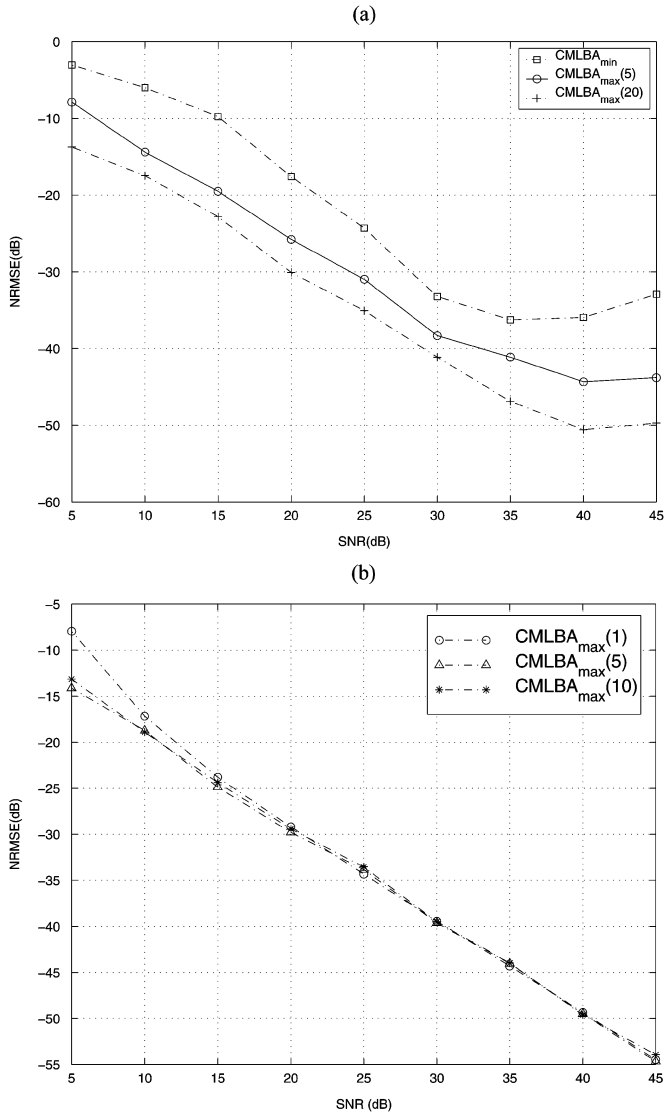


Fig. 5. Performance comparison. (a) $\mathbf{h}^{\mathrm{Hua}}$, 100 Monte Carlo runs, $M+N=58$ BPSK input symbols. (b) $\mathbf{h}^{\mathrm{Xu}}$, 100 Monte Carlo runs, $M+N=50$ QPSK input symbols.

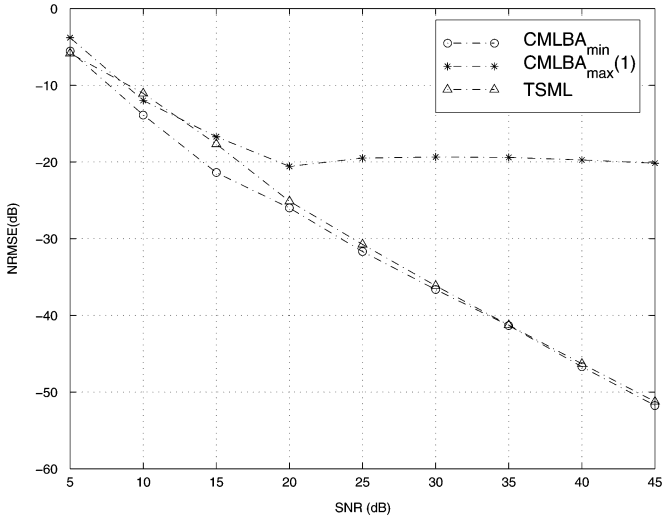


Fig. 7. Performance comparison. $\mathbf{h}^{\mathrm{Hua}}$, 100 Monte Carlo runs, $M+N=58$ 8-QAM input symbols.

- $\mathbf{h}^{\mathbf{Hua}}$: This two-channel system was first used by Hua [6]. The corresponding channel response is given by

$$\tilde{\mathbf{h}}_1^{\mathbf{Hua}}(z)=(1-e^{j\theta_1}z^{-1})(1-e^{-j\theta_1}z^{-1})$$
$$\tilde{\mathbf{h}}_2^{\mathbf{Hua}}(z)=\left(1-e^{j(\theta_1+\delta)}z^{-1}\right)\left(1-e^{-j(\theta_1+\delta)}z^{-1}\right) \quad (35)$$

where θ_1 and $\theta_1+\delta$ represent the angular position of zeros on the unit circle and δ is the distance between the zeros of the two channels. We choose to use this channel since it permits to evaluate the influence of the channel diversity. Secondly, Hua also used this channel to compare the performance of the TSML against the performance of the cross-relation algorithm [32] and of the subspace algorithm [33], and he compared both algorithms to the Cramer–Rao bound. Moreover, Zhao also used this channel to evaluate the performance of the adaptive least squares smoothing algorithm [34]. Then, by using the experimental conditions in [34], we make a fair comparison with existing approaches.

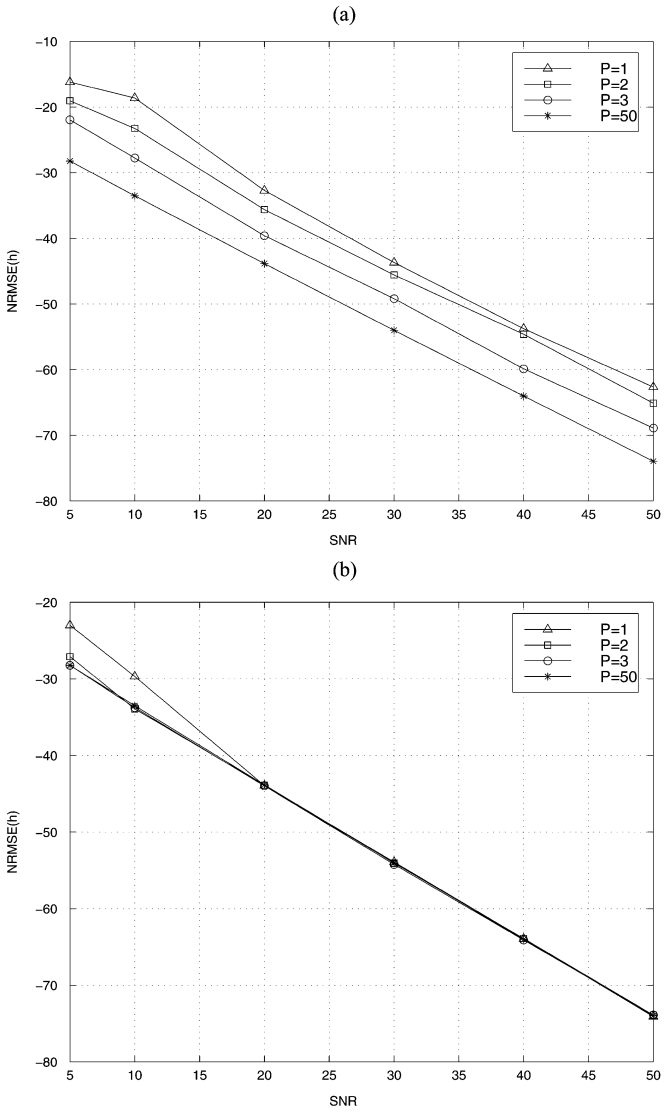


Fig. 8. NRMSE($\mathbf{h}$) versus SNR obtained with (a) CMLAA$_{\min}$ and with (b) CMLAA$_{\max}$ for various values of P ($\lambda = 1$), $N = 50$.

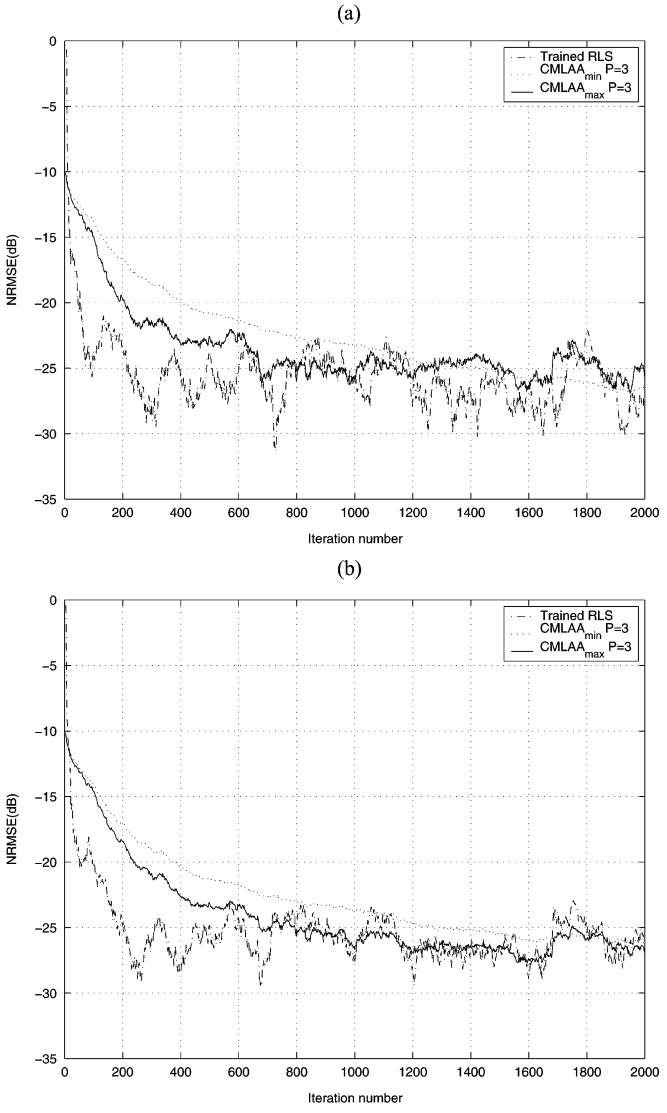


Fig. 9. NRMSE($\mathbf{h}$) versus the iteration number. SNR $= 10$ dB. Channel order (a) correctly estimated and (b) overestimated.

- $\mathbf{h^{Xu}}$: The channel responses values are given in Table I. This set of channels was first used in [35]. It considers a two-ray multipath model with delay at 0 and 1.1 baud periods. The channel model $\mathbf{h^{Xu}}$ simulates a wireless environment with long delay multipath.
- $\mathbf{h^{Tong}}$: The channel response values are given in Table II. The channel $\mathbf{h^{Tong}}$ was first used in [31]. This channel has severe intersymbol interference. It is also close to violate the identifiability condition. Moreover, $\mathbf{h^{Tong}}$ has small head and tails taps. This channel is used to test the robustness of our algorithms against the overestimation of the channel order.

2) Comments and Observations:

- Fig. 4 plots the $\mathrm{NRMSE}_{\mathrm{dB}}(h)$ against δ (relative positions of the zeros), the SNR is set to 45 dB. This SNR is quite unrealistic; however, this choice permits a fair comparison between our algorithms and TSML since TSML is biased at low SNR whereas MLBA-like algorithms are not. We also compare these methods with the MLPA [30]. We observe the following.
  - — The TSML and the MLBA have comparable performance (NRMSE≈ -50 dB when $\delta > 0.7$ and poor performance when $\delta < 0.7$).
  - — The MLPA is a generalization of the linear prediction algorithm. The MLPA exploits the channel structure completely and provides more statistical efficiency in channel identification (compared to classical LPA [28], [36]). This method has also been developed as the outer-product decomposition algorithm [37], [38]. It was also extended by Tugnait *et al.* for multiple-input multiple-output [39]. The MLPA presents poor results compared to the other methods. Actually, the simulation was run with $M + N = 32$ input symbols, whereas the MLPA exhibits good results with hundreds of input symbols. For short-burst applications, LPA-like algorithms cannot be used.
  - — The CMLBA$_{\min}$ outperforms TSML and MLBA for poor diversity conditions. Remember that CMLBA$_{\min}$

TABLE I
CHANNEL RESPONSE OF $\mathbf{h}^{\mathrm{Xu}}$

i	$\mathbf{h}_i^{\mathbf{Xu}}(0)$	$\mathbf{h}_i^{\mathbf{Xu}}(1)$	$\mathbf{h}_i^{\mathbf{Xu}}(2)$	$\mathbf{h}_i^{\mathbf{Xu}}(3)$
1	0	-1.280 - 0.301j	1.617 + 2.385j	0.178 + 0.263j
2	-1.023 - 0.501j	0.106 + 1.164j	1.477+1.850j	-0.482 - 0.523j
3	0	-0.282 + 0.562j	0.371 - 1.001j	0.041 - 0.110j
4	-0.227 + 0.487j	0.031 - 0.211j	0.336 - 0.866j	-0.110 + 0.271j

TABLE II
CHANNEL RESPONSE OF $\mathbf{h}^{\mathrm{Tong}}$

i	$\mathbf{h}_i^{\mathbf{Tong}}(0)$	$\mathbf{h}_i^{\mathbf{Tong}}(1)$	$\mathbf{h}_i^{\mathbf{Tong}}(2)$	$\mathbf{h}_i^{\mathbf{Tong}}(3)$	$\mathbf{h}_i^{\mathbf{Tong}}(4)$	$\mathbf{h}_i^{\mathbf{Tong}}(5)$
1	-0.0031 - 0.0017j	-0.0109 - 0.0025j	0.1522 + 0.0705j	0.3789 + 0.5930j	-0.0301 - 0.0348j	-0.0032 - 0.0017j
2	-0.0016 - 0.0047j	-0.0263 - 0.0433j	0.4409 + 0.4736j	0.0766 + 0.2168j	-0.0042 - 0.0154j	-0.0017 - 0.0044j

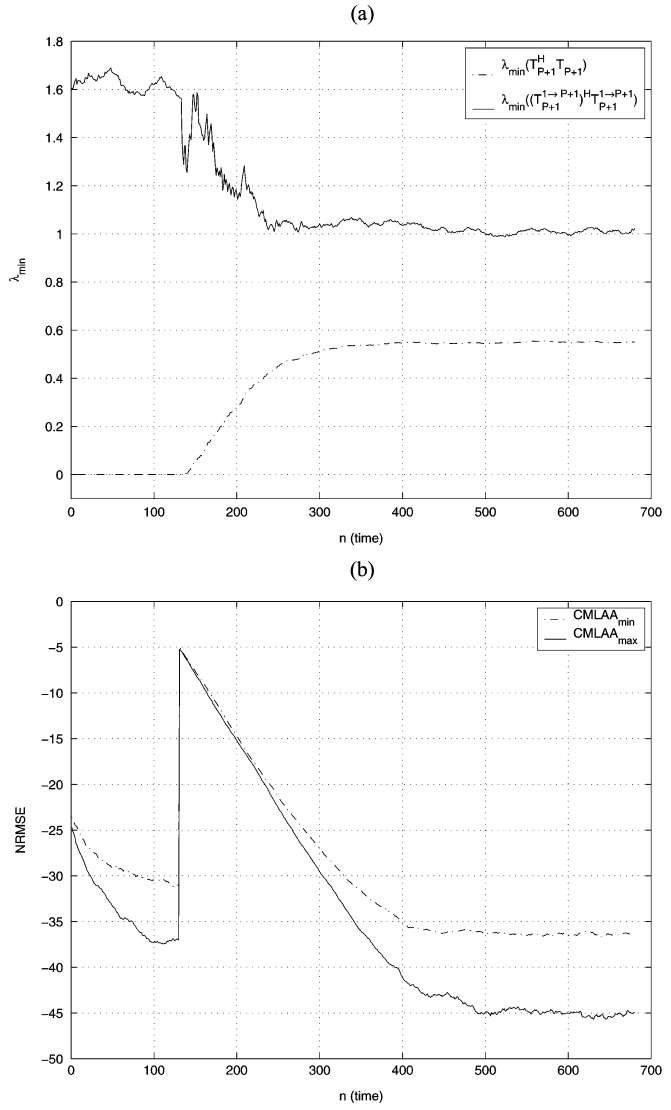


Fig. 10. Channel tracking performance (SNR = 30 dB, 20 Monte Carlo runs).

has the same local minima as MLBA. Thus this improvement (in terms of estimation accuracy) is not counterbalanced by extra local minima.

— The NRMSE of $\mathrm{CMLBA}_{\max}$ is about 5 dB less than the NRMSE of $\mathrm{CMLBA}_{\min}$, of MLBA, and of TSML when δ is large enough (thanks to the prior introduced by γ). When $\delta \to 0$, then $\gamma = \lambda_{\min} \to 0$ and $\mathrm{CMLBA}_{\max} \to \mathrm{CMLBA}_{\min}$. We are not surprised to see that the NRMSE of $\mathrm{CMLBA}_{\max}$ gets closer to the NRMSE of $\mathrm{CMLBA}_{\min}$ when δ tends toward zero.

— $\mathrm{CMLBA}_{\max}$ (5) was built to achieve good performance even when the channels share common zeros. We observe that the NRMSE of the $\mathrm{CMLBA}_{\max}$ (5) is about -55 dB whatever δ may be.

- Fig. 5 shows a comparison of the methods in various environments. In (a), $\mathbf{h}^{\mathbf{Hua}}$ is considered with $\delta = \pi/100$ and $\theta = 0$. For such a channel the smallest eigenvalue of $\mathcal{T}_N(\tilde{\mathbf{h}})^H \mathcal{T}_N(\tilde{\mathbf{h}})$ is close to zero; then $\mathrm{CMLBA}_{\min}$ and $\mathrm{CMLBA}_{\max}$ are identical and the prior knowledge is lost (we do not plot the curve corresponding to $\mathrm{CMLBA}_{\max}$). First, we can remark that the constraint of $\mathrm{CMLBA}_{\min}$ yields robustness to the lack of channel diversity. It is interesting to note that this improvement arises without increasing the number of local minima (see Proposition 1). The minimal eigenvalue is likely to increase when the partition grows. As expected, $\mathrm{CMLBA}_{\max}$ (20) outperforms $\mathrm{CMLBA}_{\max}$ (5) and $\mathrm{CMLBA}_{\min}$ even at low SNR.

  We repeat the above simulation using the set of channels $\mathbf{h}^{\mathbf{Xu}}$. This time, the smallest eigenvalue of $\mathcal{T}_N(\tilde{\mathbf{h}})^H \mathcal{T}_N(\tilde{\mathbf{h}})$ is 0.2737. Thus, in $\mathrm{CMLBA}_{\max}$ (1), the value of $\lambda_{\min}$ is no longer negligible, which leads to very good estimates of the channel. In such an environment, the partitions do not bring improvements.
- Fig. 6 shows robustness to the overestimation of the channel order and to severe intersymbol interference. The channel used for this simulation is $\mathbf{h}^{\mathbf{Tong}}$; it presents severe interference intersymbol. Moreover, the first and last taps of each subchannel have very small amplitude. We choose to run the $\mathrm{CMLBA}_{\max}$ with $M = 5$ to test the robustness of the method toward an overestimation of the channel order. The experimental conditions are identical to those in [31]. As can be seen, the method presents good results and outperforms J-LSS. The channel $\mathbf{h}^{\mathrm{Tong}}$

TABLE III
PATH PROFILE FOR CHANNEL EQUALIZATION TESTS (COST-GSM MODEL)

Path	1	2	3	4	5	6	7	8	9	10	11	12
Delay (μs)	0	0.2	0.4	0.6	0.8	1.2	1.4	1.8	2.4	3	3.2	5
Attenuation (dB)	-4	-3	0	-2	-3	-5	-7	-5	-6	-9	-11	-10

is ill conditioned; thus the quality of the estimation is improved by the partitions.

- Fig. 7 shows the behavior of the method with a non-PSK constellation. Throughout this paper, we have supposed that the emitted symbols belong to a PSK modulation H4). This assumption was at the origin of the prior knowledge introduced in the proposed criterion. The efficiency of the prior (for PSK constellations) was exhibited in the previous simulations, but what happens when a non-PSK constellation is used? Here, the input symbols belong to an 8-QAM constellation. We use $\mathbf{h}^{\mathbf{Hua}}$ with $\delta = 3\pi/8$ and $\theta = 0, M + N = 58$. Not surprisingly, it appears that $\text{CMLBA}_{\max}$ is not convenient for 8-QAM constellations. TSML and $\text{CMLBA}_{\min}$ present similar performances. However, (4) suggests that $\text{CMLBA}_{\min}$ is better to use especially for ill conditioned channels.

B. Adaptive Algorithms

In order to check the relevance of the approximations used to derive the CMLAA, we analyze the choice of the parameter P (number of symbols updated at each iteration). Then, we focus on the applicability of the proposed algorithm first by testing its robustness to an overestimation of the channel order (Fig. 9) and second by evaluating its parameter tracking performance (Fig. 10).

Fig. 8 shows the choice of P. In this simulation, the NRMSE is computed for $\text{CMLAA}_{\min}$ and for $\text{CMLAA}_{\max}$ (with $\lambda = 1$) for various pairs (P, SNR). The NRMSE is averaged over 50 Monte Carlo runs and is computed at the one-thousandth iteration. In (b), the value of the mean squared error obtained with $P = 50$ remains very close to the value obtained with $P = 3$. In (a), we observe that replacing $P = 3$ by $P = 50$ leads to the same improvement as replacing $P = 1$ by $P = 3$. Therefore, the accuracy of the channel estimate seems to be mainly influenced by the symbols in the delay line of the channels. Moreover, the computational cost of the method increases with P. In our simulation, the order of the channel is $M = 3$. Choosing P equal to the channel order appears to be a good compromise.

Fig. 9 shows robustness to channel overestimation. A 12-path propagation channel is considered, simulated according to the model of Clarke [40]. The path profile is shown in Table III. For this simulation, the channel order is set to $M = 3$, there are two subchannels, and the modulation is BPSK. The SNR, this time, is set to a more reasonable value of 10 dB.

We present the NRMSE versus the iteration number for $\text{CMLAA}_{\min}$, for $\text{CMLAA}_{\max}$, and for trained RLS. Fig. 9(a) shows the results obtained when the channel order is correctly estimated ($\hat{M} = 3$). From this figure, we can see that $\text{CMLAA}_{\max}$ outperforms $\text{CMLAA}_{\min}$ even if the algorithm does not perform a minimization with respect to the joint variable in each iteration (see Section VI-B). We can notice that the behavior of both algorithms could be improved by iterating the minimization with respect to each variable. Fig. 9(b) shows the results obtained when the channel order is overestimated ($\hat{M} = 4$). Both $\text{CMLAA}_{\min}$ and $\text{CMLAA}_{\max}$ appear to be robust to the overestimation of the channel order. This is due to the constraint introduced into the criterion.

Fig. 10 shows tracking of the channel parameters. The simulation presented here has first been experimented with by Zhao in [34]. It concerns the case where the channel order and the channel parameters have a sudden change. The initial channel is the one used for testing the batch algorithms (35) with $\delta = \pi$ and $\theta_1 = \pi/10$. The channel order and the channel parameters change at time $n = 151$ where we add zeros $Z_1 = 1$ and $Z_2 = -1$ to the two subchannels, respectively. In the simulation, the estimated channel order is fixed to $\hat{M} = 3$. The NRMSE convergence of $\text{CMLAA}_{\min}$ and of $\text{CMLAA}_{\max}$ is shown in Fig. 10(b), where we can see the ability of both algorithms to track the channel variations. In (a), the smallest eigenvalues of the full matrix $(\lambda_{\min}(\mathcal{T}_{P+1}^H \mathcal{T}_{P+1}))$ and of the truncated matrix $(\lambda_{\min}((\mathcal{T}_{P+1}^{1\to P+1})^H \mathcal{T}_{P+1}^{1\to P+1}))$ are plotted. When $n < 151$, we try to estimate a channel of order 3, whereas the order of the channel to be estimated is 2; that is the reason why $\lambda_{\min}(\mathcal{T}_{P+1}^H \mathcal{T}_{P+1}) \approx 0$. In $\text{CMLAA}_{\max}$, γ is equal to $\lambda_{\min}((\mathcal{T}_{P+1}^{1\to P+1})^H \mathcal{T}_{P+1}^{1\to P+1})$, which is not null. $\text{CMLAA}_{\max}$ takes more advantage of the prior information than $\text{CMLAA}_{\min}$ even when the subchannels share common zeros.

VIII. CONCLUSION

In this paper, a maximum likelihood approach to solve the joint blind channel identification and blind symbol estimation problem was presented. We demonstrated the improvement of the estimation accuracy by the use of a prior knowledge. Moreover, the proposed batch algorithm presents the finite-sample convergence property.

Based on this block algorithm, an adaptive version is derived by exploiting the recursive procedure proposed for solving the local minima problem. A nice advantage inherent to the use of the prior is that it brings robustness to the overestimation of the channel order thus our method does not require the channel order to be known or well estimated. Thanks to the forgetting factor, the algorithm is able to track efficiently the changes in the channel parameters. At each iteration, the number of symbols to be updated is limited to the length of the channel. Moreover, the update of the filters can be performed by stochastic gradient

techniques as shown in [10], which renders the CMLAA computationally nonexpensive. Current work on the application of these algorithms under practical situations is currently undertaken, and will be reported.

REFERENCES

[1] Z. Ding and Y. Li, *Blind Equalization and Identification*, ser. Signal Processing and Communications Series: Marcel Dekker, 2001.

[2] J. K. Tugnait, "Blind equalization techniques," in *Encyclopedia of Telecommunications*. New York: Wiley, 2002.

[3] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, "Channel estimation and equalization using higher-order statistics," in *Signal Processing Advances in Wireless and Mobile Communications: Trends in Channel Estimation and Equalization*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[4] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf. Signals, Systems, and Computers*, Nov. 1991, pp. 856–860.

[5] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, 1998.

[6] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, pp. 661–672, Mar. 1996.

[7] D. Slock and C. B. Papadias, "Further results on blind identification and equalization of multiple FIR channels," in *Proc. ICASSP*, 1995, pp. 1964–1967.

[8] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery-application to equalization in underwater communications," *IEEE J. Ocean. Eng.*, vol. 16, pp. 42–55, Jan. 1991.

[9] D. Gesbert, P. Duhamel, and S. Mayrargue, "Blind least-squares approaches for joint data/channel estimation," in *IEEE DSP Workshop*, Sept. 1996.

[10] F. Alberge, P. Duhamel, and M. Nikolova, "Adaptive solution for blind identification/equalization using deterministic maximum likelihood," *IEEE Trans. Signal Process.*, vol. 50, pp. 923–936, Apr. 2002.

[11] N. Seshadri, "Joint data and channel estimation using fast blind treillis search techniques," in *Proc. Globecom*, 1990, pp. 1659–1663.

[12] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," *Opt. Eng.*, vol. 31, no. 6, pp. 1224–1228, Jun. 1992.

[13] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Process. Lett.*, vol. 1, pp. 29–31, Feb. 1994.

[14] M. Nikolova, "Estimation of binary images by minimizing convex criteria," in *Proc IEEE Int. Conf. Image Processing*, vol. 2, Oct. 1998, pp. 108–112.

[15] W. K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P. C. Ching, "Quasi maximum likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, 2002.

[16] P. H. Tan, L. K. Rasmussen, and T. J. Lim, "Constrained maximum likelihood detection in CDMA," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 142–153, 2001.

[17] R. E. Blahut, *Algebraic Methods for Signal Processing and Communication Coding*. New York: Springer-Verlag, 1992.

[18] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[19] E. Pité and P. Duhamel, "Bilinear methods for blind channel equalization: (no) local minimum issue," in *Proc. ICASSP*, May 1998, pp. 2113–2116.

[20] C. B. Papadias, "Methods for blind equalization and identification of linear channels," Ph.D. dissertation, ENST, Paris, France, 1995.

[21] K. G. Murty and S. N. Kabadi, "Some NP-complete problems in quadratic and nonlinear programming," *Math. Program.*, vol. 39, pp. 117–129, 1987.

[22] P. M. Pardalos and G. Schnitger, "Checking local optimality in constrained quadratic programming is np-hard," *OR Lett.*, vol. 7, pp. 33–35, 1988.

[23] R. Horst and H. Tuy, *Global Optimization (Deterministic Approaches)*. Berlin, Germany: Springer-Verlag, 1993.

[24] I. Bomze and G. Danninger, "A finite algorithm for solving general quadratic problem," *J. Global Opt.*, vol. 4, pp. 1–16, 1994.

[25] O. Barrientos and R. Correa, "An algorithm for global minimization of linearly constrained quadratic functions," *J. Global Opt.*, vol. 16, pp. 77–93, 2000.

[26] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[27] P. G. Ciarlet, *Introduction to Numerical Linear Algebra and Optimization*: Cambridge Univ. Press, 1989.

[28] D. Slock, "Blind fractionaly-spaced equalization, perfect reconstruction filter banks, and multichannel linear prediction," in *Proc. ICASSP*, 1994, pp. 585–588.

[29] E. de Carvalho, "Identification de canal et egalization aveugles et semi-aveugles pour les communications mobiles," Ph.D. dissertation, ENST, Paris, France, 1999.

[30] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. ICASSP*, Apr. 1997, pp. 3621–3624.

[31] L. Tong and Q. Zhao, "Joint order detection and blind channel estimation by least squares smoothing," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2345–2355, 1999.

[32] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind channel identification of multi-channel FIR systems," in *Proc. ICASSP*, Apr. 1994.

[33] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, pp. 516–526, Feb. 1995.

[34] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Process.*, vol. 47, pp. 3000–3012, Nov. 1999.

[35] G. Xu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, 1995.

[36] K. A. Meraim, P. Duhamel, D. Gesbert, and P. Loubaton, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, May 1995.

[37] Z. Ding, "An outer-product decomposition algorithm for multichannel blind identification," in *Proc. 8th IEEE Workshop Stat. Signal Array Processing*, June 1996, pp. 132–135.

[38] ——, "Matrix outer-product decomposition method for blind multiplechannel identification," *IEEE Trans. Signal Process.*, vol. 45, pp. 3053–3061, Dec. 1997.

[39] J. K. Tugnait and B. Huang, "Multistep linear predictors-based blind identification and equalization of multiple-input multiple output channels," *IEEE Trans. Signal Process.*, vol. 48, pp. 26–38, Jan. 2000.

[40] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, 1968.

Florence Alberge was born in Albi, France, in 1971. She received the engineering degree from Ecole Nationale Supérieure de l'Electronique et de ses Applications, France, and the master's degree from the University of Cergy, France, in 1996. She received the Agregation de Physique and Ph.D. degrees from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1998 and 1999, respectively.

Since 2000, she has been with the Laboratoire des Signaux et Systèmes, Gif sur Yvette, France, and with the University of Paris-Sud, Orsay, France, as an Assistant Professor. Her main research interests are in the field of signal processing for communications.

Mila Nikolova is a Researcher with the National Center for Scientific Research (CNRS), France. She is also with the Center for Mathematics and their Applications (CMLA), ENS de Cachan, France. Her research interests include inverse problems, mathematical image and signal processing, and variational problems and their analysis.

Pierre Duhamel (SM'87–F'98) was born in France in 1953. He received the Eng. degree in electrical engineering from the National Institute for Applied Sciences (INSA) Rennes, France, in 1975 and the Dr.Eng. and Doctorat ès Sciences degrees from Orsay University, Orsay, France, in 1978 and 1986, respectively.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive charge-coupled device filters. Later, he worked on fast algorithms for computing Fourier transforms and convolutions and applied similar techniques to adaptive filtering, spectral analysis, and wavelet transforms. From 1993 to 2000, he was a Professor with Ecole Nationale Supérieure des Télécommunications, Paris, with research activities focused on signal processing for communications. He was Head of the Signal and Image Processing Department from 1997 to 2000. He is now with the Laboratoire de Signaux et Systemes, Gif sur Yvette, France, where he is developing studies in signal processing for communications (including equalization, iterative decoding, multicarrier systems) and signal/image processing for multimedia applications, including source coding, joint source/channel coding, watermarking, and audio processing.

Dr. Duhamel was Co-General Chair of the 2001 International Workshop on Multimedia Signal Processing, Cannes, France, and will be Co-Technical Chair of ICASSP'06, Toulouse, France. He was Chairman of the IEEE DSP Committee from 1996 to 1998 and a member of the SP for Com Committee until 2001. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, an Associate Editor of IEEE SIGNAL PROCESSING LETTERS, and a Guest Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING Special Issue on Wavelets. He was an IEEE Distiguished Lecturer for 1999. He was a Corecipient of the Best paper Award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1998. He received the Grand Prix France Telecom award from the French Science Academy in 2000.

Annexe 3

Z. Mheich, F. Alberge, P. Duhamel. Achievable rates optimization for broadcast channels using finite size constellations under transmission constraints. EURASIP Journal on Wireless Communications and Networking, 2013.

Abstract

In this paper, maximal achievable rate regions are derived for power-constrained AWGN broadcast channel involving finite constellations and two users. The achievable rate region is studied for various transmission strategies including superposition coding and compared to standard schemes such as time sharing. The maximal achievable rates are obtained by optimizing over both the joint distribution of probability and over the constellation symbol positions. A numerical solution is proposed for solving this non-convex optimization problem. Then, we consider several variations of the same problem by introducing various constraints on the optimization variables. The aim is to evaluate efficiency vs. complexity tradeoffs of several transmission strategies, some of which (the simplest ones) can be found in actual standards. The improvement for each scheme is evaluated in terms of SNR savings for target achievable rates or/and percentage of gain in achievable rates for one user compared to a reference scheme. As an application, two scenarios of coverage areas and user alphabets are considered. This study allows to evaluate with practical criteria the performance improvement brought by more advanced schemes.

EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Achievable rates optimization for broadcast channels using finite size constellations under transmission constraints

Zeina Mheich[1,2,3], Florence Alberge[1,2,3*] and Pierre Duhamel[1,2,3]

Abstract

In this paper, maximal achievable rate regions are derived for power-constrained AWGN broadcast channel involving finite constellations and two users. The achievable rate region is studied for various transmission strategies including superposition coding and compared to standard schemes such as time sharing. The maximal achievable rates are obtained by optimizing over both the joint distribution of probability and over the constellation symbol positions. A numerical solution is proposed for solving this non-convex optimization problem. Then, we consider several variations of the same problem by introducing various constraints on the optimization variables. The aim is to evaluate efficiency vs. complexity tradeoffs of several transmission strategies, some of which (the simplest ones) can be found in actual standards. The improvement for each scheme is evaluated in terms of SNR savings for target achievable rates or/and percentage of gain in achievable rates for one user compared to a reference scheme. As an application, two scenarios of coverage areas and user alphabets are considered. This study allows to evaluate with practical criteria the performance improvement brought by more advanced schemes.

Keywords: AWGN broadcast channels; Achievable rate region; Hierarchical modulation; Superposition modulation; Superposition coding; Constellation shaping; Non-convex optimization

1 Introduction

During the past few decades, information networks have witnessed tremendous and rapid advances, based on the important growth in the adoption of new wireless technologies, applications and services, first from cellular networks and more recently for computer networks (WLANs). Consequently, wireless networks are exposed to capacity and coverage problems, and the focus is now shifting towards capturing some of the aspects of realistic networks by studying natural network models such as models with broadcasting.

In 1972, achievable rate region is obtained by Cover in [1] for Gaussian broadcast channels with two outputs and generalized by Bergmans to broadcast channels with any number of outputs [2]. Roughly a year later, the optimality of the sets of achievable rates was established by Bergmans [3] and Gallager [4]. Superposition coding is a possible solution to achieve good rate regions in which information intended for high-noise receivers and information intended for low-noise receivers are superimposed and transmitted simultaneously on the same radio resource. The low-noise receivers can always decode messages intended for the high-noise receivers. Thus, they effectively cancel out the interference due to the signal intended for the high-noise receivers, and then decode their own message. The high-noise receivers decode their messages by treating the low-noise receivers message as noise. Superposition coding appears in several contexts in information theory and is closely related to multilevel coding and unequal error protection [5,6]. Cover showed [1] that the superposition coding reaches the theoretical limit of the capacity region for two user Gaussian broadcast channel using an infinite Gaussian input alphabet for each user. A treatment of the case of

*Correspondence: alberge@lss.supelec.fr
[1]University Paris-Sud, UMR8506 Orsay, F-91405, France
[2]CNRS, Gif-sur-Yvette, F-91192, France
Full list of author information is available at the end of the article

multiple transmitter/receivers for the band-limited additive white Gaussian noise channel is given by Bergmans and Cover in [7], where it is proved that superposition coding can achieve higher-rate region than orthogonal schemes such as frequency-division multiple access (FDMA) or time division multiple access (TDMA). However, in actual transmission systems, the channel input is constrained to a finite size alphabet with equal probability symbols. A well-known practical implementation of superposition coding is hierarchical modulation, also called layered modulation, which uses constellations with non-uniformly spaced signal points creating different levels of error protection. Hierarchical modulation is used to mitigate the cliff effect in digital television broadcast and is included in various standards, such as Digital Video Broadcast for Terrestrial Television (DVB-T) [8], DVB to Handhelds (DVB-H), and DVB Satellite services to Handhelds (DVB-SH) [9] standard proposal for mobile digital TV transmission. A study about the performance of hierarchical modulation and a comparison with time sharing strategy in terms of achievable rates can be found in [10].

The restriction imposed by practical systems in using finite signaling constellation and equiprobable symbols reduces the achievable rates and leads to a gap with the capacity region achieved with Gaussian input alphabets for AWGN broadcast channel. This gap can be reduced using a technique called constellation shaping. In fact, most results for constellation shaping with finite signal constellations consider only point-to-point communication systems [11]. Then, the concept of constellation shaping has been adapted to most modern coding and modulation techniques as for example turbo coding and BICM schemes [12-19]. For broadcast channels, the achievable rate region for two-user AWGN broadcast channels with finite input alphabets is derived in [20] when superposition of modulated signal is used as transmission strategy. In their work, the authors assume a uniform distribution over the finite input set. To our knowledge, no study is available about the maximization of the achievable rate region for two-user AWGN broadcast channels with finite size constellations by optimizing over both the joint probability distribution and constellation symbol positions for a broadcast transmission strategy. This general framework encompasses hierarchical modulations as a special case. In this paper, maximal achievable rate regions are derived for power-constrained AWGN broadcast channel of two users with M-pulse amplitude modulation (M-PAM) constellations of M points using various transmission strategies. A numerical solution is proposed for solving this non-concave optimization problem. In a typical broadcast system, there is a trade off between achievable rates and coverage areas. Therefore, we are interested in determining the transmission strategy which provides the best achievable rates or the maximal SNR gain for a given coverage scenario. The compromise between the simplicity of implementation and expected gains is also evaluated.

The organization of the paper is as follows. Section 2 recalls some information theory results on broadcast channels and degraded broadcast channels. In section 3, various transmission strategies for broadcast systems are described. Section 4 gives a formulation of the problem in terms of optimization for the various transmission strategies under consideration. Then, computational aspects are discussed. An iterative algorithm is proposed for the computation of maximal achievable rate regions using superposition coding (general case) and M-PAM constellation or in the particular case of superposition modulation. The proposed algorithm can handle an optimization with respect to the joint distribution of probability or with respect to the positions of constellation symbols. Both variables can also be considered jointly. Obviously, the best results are obtained for the most general case. Our target is to (1) evaluate the loss experienced using simple schemes, (2) identify situations in which complex schemes (non-standard) lead to significant improvements. As an application, we consider, in section 5, several scenarios of coverage areas and user alphabets, and we give conclusions about the transmission strategies which can provide the best trade off between efficiency and complexity of implementation.

2 AWGN broadcast channels

A two-receiver (users) broadcast channel (BC) consists of an input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}_1$ (user 1), $\mathcal{Y}_2$ (user 2), and a conditional pdf $P_{Y_1Y_2|X}$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$. Let X, Y_1, and Y_2 be random variables representing the input and outputs of the BC. Figure 1 depicts the two users BC with two independent messages W_1 and W_2. The encoder generates a codeword $x^n(w_1, w_2)$ of length n based on these two messages. Each user receives, respectively, y_1^n and y_2^n. A BC is said to be physically degraded if $P_{Y_1Y_2|X}(y_1, y_2|x) = P_{Y_1|X}(y_1|x) \cdot P_{Y_2|Y_1}(y_2|y_1)$ (i.e., $X \rightarrow Y_1 \rightarrow Y_2$ form a Markov chain). A BC is said to be stochastically degraded or degraded if there exists a random variable $\tilde{Y}_1$ which has the same conditional pdf as Y_1 given X, such that $X \rightarrow \tilde{Y}_1 \rightarrow Y_2$ forms a Markov chain. We are interested in degraded BC because its capacity

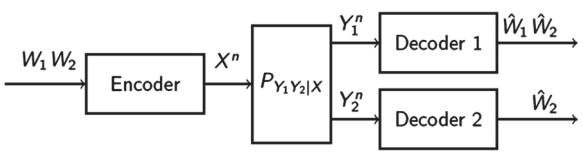


Figure 1 The two-user broadcast channel.

region is known, while it is not available for the general case.

In our system model, W_1 denotes the private message intended for receiver 1 only, and W_2 is a common message for both receivers. A typical example of this situation is digital TV broadcasting to two different groups of receivers, classified according to their channel conditions, where the basic signal (common signal) should be available to all receivers. The higher quality is realized by adding the basic signal with an incremental signal (private signal for receivers of good channel conditions) which carries TV signal with a high data rate, such as HDTV.

Let R_1 and R_2 be the rates at which the transmitter is sending W_1 and W_2, respectively. Thus, user 1 achieves $R_1 + R_2$, while user 2 achieves R_2. The capacity region of the degraded broadcast channel $X \rightarrow Y_1 \rightarrow Y_2$ in Figure 1 is the convex hull of the closure of rate pairs $(R_1 + R_2, R_2)$ satisfying

$$R_1 \leq I(X; Y_1|U) \tag{1}$$

$$R_2 \leq I(U; Y_2) \tag{2}$$

for some joint distribution $P_{UXY_1Y_2} = P_{UX} \cdot P_{Y_1|X} \cdot P_{Y_2|X}$ on $\{\mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2\}$ [21]. $P_{Y_1|X}$ and $P_{Y_2|X}$ are conditional pdfs that depend on the channel model. P_{UX} is the joint probability distribution of U and X, where the auxiliary random variable U has cardinality bounded by $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$. The capacity region is achieved using superposition coding, where U serves as the center of a cloud of codewords that can be distinguished by both receivers. Since the capacity region of a BC depends only on the conditional marginals, the capacity region of the stochastically degraded BC is equal to that of the corresponding physically degraded channel. Cover [1] showed that in the case of binary symmetric BC and AWGN BC, superposition coding expands the rate region beyond that achievable with time sharing.

Now, consider the Gaussian broadcast channel with two users. Without loss of generality, assume that Y_1 is less noisy than Y_2. It can easily be shown that scalar Gaussian broadcast channels are equivalent to a degraded channel,

$$Y_1 = X + Z_1 \tag{3}$$

$$Y_2 = X + Z_2 = Y_1 + Z_2', \tag{4}$$

where $Z_1 \sim \mathcal{N}(0, \sigma_1^2), Z_2 \sim \mathcal{N}(0, \sigma_2^2), Z_2' \sim \mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$, and Z_1, Z_2' are independent. Thus, Gaussian BC is stochastically degraded. We assume an average power constraint on the transmitted power P defined as $\mathbb{E}[X^2] \leq P$. The received signal to noise ratio for each user is $\mathrm{SNR}_i = \frac{P}{\sigma_i^2}$, where $\mathrm{SNR}_1 > \mathrm{SNR}_2$, and σ_i^2 is the variance of the noise Z_i. The capacity region of the AWGN-BC is the set of rate pairs $(R_1 + R_2, R_2)$, such that

$$R_1 \leq C(\alpha \cdot \mathrm{SNR}_1) \tag{5}$$

$$R_2 \leq C\left(\frac{(1-\alpha) \cdot \mathrm{SNR}_2}{\alpha \cdot \mathrm{SNR}_2 + 1}\right) \tag{6}$$

for all $\alpha \in [0, 1]$, where $C(x) = \frac{1}{2} \cdot \log_2(1+x)$. The theoretical limit of two-user AWGN BC is achieved using signal superposition [1].

3 Broadcast transmission strategies

In this section, various transmission strategies for broadcast systems are described. The strategies are presented in ascending order of implementation complexity. Specifically, by moving from one strategy to another, we release some constraints on the system implementation to reach finally the most complex strategy that can be used to broadcast information for users. Obviously, since the simple schemes can be understood as adding constraints to the most general case, they are less efficient in terms of attainable rates.

3.1 Time sharing

Time sharing (TS) has been widely used in broadcast systems as broadcast transmission strategy. In time sharing scheme, a percentage of time is used to send one message, and the rest of the time is used to send another message. Thus, it is practical to implement because the rate pairs can be achieved by strategies used for point-to-point channel and sharing the time between messages. As in previous works on broadcasting, this situation serves as a reference for the more advanced schemes.

In this work, a time sharing scheme with standard constellation M-PAM (Figure 2) is considered when symbols are used with equal probability. A standard M-PAM constellation is defined as a constellation with M real symbols belonging to $\mathcal{X} = \{M - 1 - 2 \cdot (i-1), \text{for } i = 1, \ldots, M\}$. During the time slot dedicated to send a message, only one data stream is sent using the entire set of constellation points. In classical implementations of time sharing,

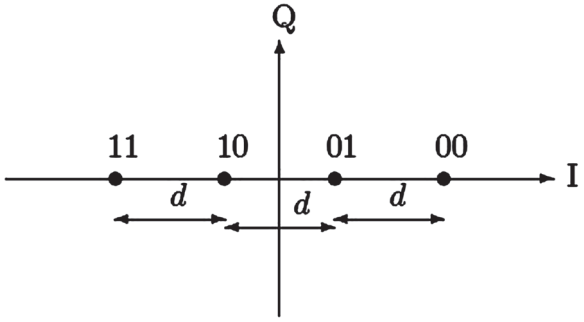


Figure 2 4-PAM with equally spaced symbols.

the conventional M-PAM symbols are equally spaced and used with equal probability.

3.2 Hierarchical Modulation (HM)

In two-layer hierarchical modulation, constellation symbols are used to transmit two data streams simultaneously for two users [22,23]. Constellation symbols are usually chosen with the same probability but may be non-equally spaced. These symbols can be considered as the sum of two lower-order modulations, one for each user. The modulation with higher power is used for the 'bad' channel, the one with smallest power for the 'good' channel. Hence, the encoding using hierarchical modulation can be separable for the two streams which is more practical.

This is explained here using 4-PAM as an example. Figure 3 shows the constellation diagram of a hierarchical 4-PAM with parameter $\ell = \ell_1/\ell_2$ used to determine the spacing between the groups of constellation points (clouds). ℓ is the ratio of the spacing between the groups to the spacing between individual points within a group. Standard values of ℓ are 1, 2, and 4. When ℓ increases, with a fixed total transmission power P, the two points from both sides of origin form a cloud. The location of a point within its cloud is regarded as the information for the 'good' user. The other information, i.e., the number of the cloud in which the point is located is the information for the 'bad' user. In this way, two separate data streams can be made available for transmission. Formally, we are still dealing with 4-PAM, but in the hierarchical interpretation, it is viewed as the combination of two BPSK modulations which have different robustness to noise. In other words, the service coverage areas differ in size for both users. The better-protected data stream is referred to as the high-priority (HP) stream which is mapped in Figure 3 to the most significant bit. The other one is referred to as the low-priority (LP) stream (Figure 3) and mapped in Figure 3 to the least significant bit. Receivers with good reception conditions can receive both streams, while those with poorer reception conditions may only receive the high priority stream considering the LP stream as noise. This corresponds to a specific labeling of the modulation.

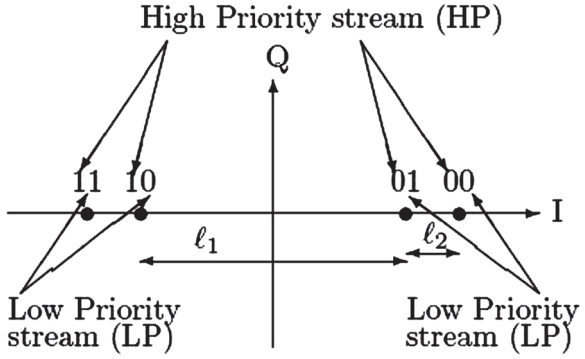


Figure 3 Hierarchical 4-PAM with parameter $\ell = \ell_1/\ell_2$.

3.3 Superposition modulation

In superposition modulation (SM) [24], the M constellation points are used such that the labeling is separable, i.e., $M = M_1 M_2$, and that the M points are obtained by adding (in $\mathbb{R}$) two rv's X_1 and X_2 of cardinality M_1 and M_2, respectively ($M_1, M_2 \in \mathbb{N} \setminus \{0,1\}$). Thus, this scheme is with an enlarged set of feasible labelings than in the previous case [25,26]. This leads also to $U \equiv X_2$ for superposition modulation because user 2 can distinguish only U.

This work studies several cases of superposition modulation. First, when the constellation symbols for each user are used with equal probability. This case will be denoted as $\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$. This is a practical case since the encoding of the messages is separable, and the symbols are used with equal probability as in real transmission systems. Then, the constraint of using equiprobable symbols is released and the symbols of user constellations can be dependent and used with non-equal probability (P_{UX} non-uniform). Thus, the encoding here is done jointly for the two messages. This strategy will be denoted $\mathrm{SM}_{\overline{\mathcal{X}},P_{UX},P_X}$ when the symbols take the values of a standard M-PAM and $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$, otherwise. In the latter case, the symbol positions can take arbitrary values and will be considered as variables to be optimized. The definition of superposition modulation can be generalized using more general form for P_{UX} than the uniform case. In superposition modulation, 2^{nR_2} independent codewords $u^n = x^{(2)n}(w_2)$ of length n are generated according to P_U; for each of these codewords, 2^{nR_1} satellite codewords $v^n = x^{(1)n}(w_1)$ are generated and added to form codewords $x^n(w_1, w_2) = u^n + v^n$ according to $P_{X|U}$. Thus, the fine information v^n is superimposed on the coarse information u^n.

Note that the capacity region of Gaussian broadcast channel is achieved using this coding scheme and successive cancellation decoding, where U ($\equiv X_2$) and V ($\equiv X_1$) are independent random variables following normal distributions. However, we do not assume here that U and V are independent. Consequently, for superposition modulation, P_{UX} takes a specific expression. As an example, consider an 8-PAM modulation. In that case, the transmitted signal at time k is the sum of the two users signals and is given by $x_k = x_k^{(1)} + x_k^{(2)}$, where $x_k^{(1)} \in \mathcal{X}_1$ and $x_k^{(2)} \in \mathcal{X}_2$ with $M_1 \cdot M_2 = 8$. Two configurations are possible either $M_2 = 4$ ($\mathcal{X}_1$ is a BPSK, and $\mathcal{X}_2$ is a 4-PAM) or $M_2 = 2$ ($\mathcal{X}_1$ is a 4-PAM, and $\mathcal{X}_2$ is a BPSK). In

both cases, P_{UX} is a sparse matrix of size $M_2 \times M$ with expression

$$P_{UX}=\begin{bmatrix} p_{00} & p_{01} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & p_{13} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{24} & p_{25} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{36} & p_{37} \end{bmatrix} \text{ if } M_1=2, M_2=4 \tag{7}$$

$$P_{UX}=\begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{14} & p_{15} & p_{16} & p_{17} \end{bmatrix} \text{ if } M_1=4, M_2=2, \tag{8}$$

where $P_{UX}[i,j] = p_{i-1,j-1} = \Pr\{U = u_{i-1}, X = x_{j-1}\}$. In both cases, the number of elements to be computed is 8.

Note also that P_{UX} and $\mathcal{X}$ (of cardinality M) determine the labeling of the input signal constellation for a fixed labeling for $\mathcal{X}_1$ and $\mathcal{X}_2$ [25,26]. Thus, the information can be distinguished using the labeling. Consider for example a label l_k^u of $log_2(|\mathcal{X}_2|)$ binary labels for u_k and l_j^v of $log_2(|\mathcal{X}_1|)$ binary labels for v_j with $k \in \{0,..,|\mathcal{X}_2|-1\}$ and $j \in \{0,..,|\mathcal{X}_1|-1\}$. Obviously, the M symbols x_i, $i \in \{0,..,|\mathcal{X}|-1\}$ carry $log_2(M)$ binary labels which are the concatenations of the labels of u_k and v_j such as $x_i = u_k + v_j$.

Part of this work on superposition modulation was presented in [25-27], where the achievable rate regions for $\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$ and $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ strategies are analyzed using a 4-PAM constellation in [25,26] and for {4,8,16}-PAM constellations in [27]. In this work, the achievable rates are also derived for $\mathrm{SM}_{\overline{\mathcal{X}},P_{UX},P_X}$ using {4,8,16}-PAM constellations.

3.4 Superposition coding

Superposition coding (SC) is one of the basics of coding schemes in network information theory. This idea was first introduced by Cover in an information theoretic study of broadcast channels [1]. In superposition coding, the joint distribution of probability P_{UX} can take a more general form than in the case of superposition modulation. In this case, the labeling cannot distinguish between the common information and the private information for user 1, a fact which increases the decoder complexity. Indeed, since the auxiliary random variable U has cardinality bounded by $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$, we use the name general superposition coding or superposition coding simply to describe the case, where $|\mathcal{U}| = \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$. For superposition coding and with M-PAM modulation, P_{UX} is an $M \times M$ matrix with elements $p_{i,j}$.

The basics of superposition coding are briefly recalled below; a detailed description is given in [28]. In this scheme, 2^{nR_2} sequences $u^n(w_2), w_2 \in [1, 2^{nR_2}]$ each i.i.d., are generated randomly and independently to represent the coarse message, each according to $\prod_{i=1}^n p_U(u_i)$. For each auxiliary sequence, $u^n(w_2)$ randomly, conditionally, and independently generates 2^{nR_1} sequences $x^n(w_1, w_2)$ and $w_1 \in [1, 2^{nR_1}]$, each according to $\prod_{i=1}^n p_{X|U}(x_i|u_i(w_2))$ to represent the fine message w_1. Thus, in superposition coding, the auxiliary random variable U serves as a cloud center for the information, distinguishable by both receivers. In this case, the decoding of information by users is based on large block joint typicality. This comes in contrast with the simpler cases where the message for user 2 was carried by the center of *modulation* clouds which imply a possible scalar detection.

The achievable rates for superposition coding will be studied for various strategies corresponding to different constraints on P_{UX} and/or $\mathcal{X}$. An exhaustive list of all the strategies under consideration is given in Table 1, where redundant configurations are omitted.

4 Achievable rate regions

For a two-user Gaussian BC, the theoretical limit of the capacity region is achieved using Gaussian input alphabet for each user. However, practical implementation constraints impose the use of finite input alphabets, and the symbols are usually chosen with equal probability. These restrictions contribute to increase the gap between the capacity region achieved with infinite Gaussian inputs and the throughput obtained in practical situations. In this section, we are interested in computing the achievable rate region of power-constrained AWGN BC when the transmitted signal is modulated using an M-PAM constellation, under the various situations described above. Since the last case (superposition coding) encompasses all previous ones as special cases, the corresponding optimization problems can be solved with the same strategy, which is detailed in this section.

Table 1 Strategies under consideration

Transmission	Variables	Constraints	Designation
SM	$\mathcal{X}$	Uniform distribution for P_{UX}	$\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$
SM	P_{UX} s.t. $\sum_{i,j} p_{i,j} = 1$	Symbol locations: M-PAM	$\mathrm{SM}_{\overline{\mathcal{X}},P_{UX},P_X}$
SM	$\mathcal{X}$		$\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$
	P_{UX} s.t. $\sum_{i,j} p_{i,j} = 1$		
SC	P_{UX} s.t. $\sum_i p_{i,j} = \frac{1}{M}$	Symbol locations: M-PAM	$\mathrm{SC}_{\overline{\mathcal{X}},P_{UX},\overline{P_X}}$
		Uniform distribution for P_X	
SC	$\mathcal{X}$	Uniform distribution for P_X	$\mathrm{SC}_{\mathcal{X},P_{UX},\overline{P_X}}$
	P_{UX} s.t. $\sum_i p_{i,j} = \frac{1}{M}$		
SC	P_{UX} s.t. $\sum_{i,j} p_{i,j} = 1$	Symbol locations: M-PAM	$\mathrm{SC}_{\overline{\mathcal{X}},P_{UX},P_X}$
SC	$\mathcal{X}$		$\mathrm{SC}_{\mathcal{X},P_{UX},P_X}$
	P_{UX} s.t. $\sum_{i,j} p_{i,j} = 1$		

4.1 Problem formulation

Consider a two-user memoryless AWGN broadcast channel ($\mathrm{SNR}_1 > \mathrm{SNR}_2$) with signal power constraint P. The channel input belongs to a finite set $\mathcal{X} = \{x_0, \ldots, x_{M-1}\} \subset \mathbb{R}$ represented by an M-PAM constellation. Assume a symmetric input signal constellation with respect to the origin. Since $\mathcal{U}$ has cardinality bounded by $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$, and the output alphabet cardinality for an AWGN channel is infinite, we have $|\mathcal{U}| \leq |\mathcal{X}|$. Thus, $|\mathcal{U}| \leq M$.

To determine the maximal achievable rate region using superposition coding, consider the case $|\mathcal{U}| = M$. For superposition modulation, we take into account the specificity on P_{UX} given in section 3.3. We also consider within the same framework the problem of maximizing the achievable rates under additional constraints on optimization variables (P_{UX} and $\mathcal{X}$): standard M-PAM symbols values, uniform distribution for P_{UX}, uniform distribution for P_X. The problem of maximizing the achievable rates under a specific situation is solved subject to a combination of constraints according to Table 1. We recall that in this work, message w_2 is a common message to both receivers, and w_1 is a private message to user 1. Thus, the achievable rate region (R_2 vs. $R_1 + R_2$) can be obtained by solving the weighted sum rate ($\theta \cdot R_1 + (1-\theta) \cdot R_2$) maximization for $\theta \in [0, 0.5]$. Indeed, for $\theta = 0$, we maximize the common information rate R_2, and when $\theta = 0.5$, we maximize the rate achieved by user 1 ($R_1 + R_2$). Using (1) and (2), the optimization problem under consideration is:

$$\begin{array}{ll} \max_{P_{UX},\mathcal{X}} & \theta \cdot I(X;Y_1|U) + (1-\theta) \cdot I(U;Y_2) \\ s.t. & p_{ij} \geq 0 \quad \forall i,j \\ & \sum_{i,j} p_{ij} \cdot x_j^2 \leq P \end{array} \tag{9}$$

and subject to the constraint on the joint pdf P_{UX} or on $\mathcal{X}$ given in Table 1 for each strategy, where $p_{ij} = \Pr\{U = u_i, X = x_i\}, j \in \{0, .., M-1\}$, and $i \in \{0, .., |\mathcal{U}| - 1\}$. The two mutual information $I(X;Y_1|U)$ and $I(U;Y_2)$ can be written as follows:

$$I(X;Y_1|U) = \sum_{i,j} \int_{-\infty}^{+\infty} p_{ij} P_{Y_1|X}(y_1|x_j) \times \log \frac{(\sum_{j'} p_{ij'}) P_{Y_1|X}(y_1|x_j)}{\sum_{j'} p_{ij'} P_{Y_1|X}(y_1|x_{j'})} dy_1 \tag{10}$$

$$I(U;Y_2) = \sum_{i} \int_{-\infty}^{+\infty} (\sum_{j} p_{ij} P_{Y_2|X}(y_2|x_j)) \times \log \frac{\sum_{j'} p_{ij'} P_{Y_2|X}(y_2|x_{j'})}{(\sum_{j'} p_{ij'})(\sum_{i',j'} p_{i'j'} P_{Y_2|X}(y_2|x_{j'}))} dy_2, \tag{11}$$

where all logarithms are taken base 2. The AWGN channel for each user is characterized by the conditional pdf

$$P_{Y_i|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} . e^{-\frac{(y-x)^2}{2\sigma_i^2}} \quad i \in \{1, 2\}. \tag{12}$$

When $\theta = 0$ or $\theta = 1$ and for $|\mathcal{U}| = M$ (which are referred in this paper as point-to-point (PtP) channel case), the individual achievable rates R_2 and R_1 are maximized respectively. The problem (9) is equivalent to

$$\begin{array}{ll} \max_{P_X,\mathcal{X}} & I(X;Y_k) \\ s.t. & p_i \geq 0 \quad \forall i \\ & \sum_i p_i = 1 \\ & \sum_i p_i \cdot x_i^2 \leq P, \end{array} \tag{13}$$

where $p_i = \Pr\{X = x_i\}, i \in \{0, .., M-1\}$ is the input probability distribution, and $k \in \{1, 2\}$. When $\theta = 0$ or 1, problem (13) is solved for $k = 2$ and 1, respectively, with $I(X;Y_k)$ given by

$$I(X;Y_k) = \int_{-\infty}^{+\infty} \sum_j p_j P_{Y_k|X}(y_k|x_j) \times \log \frac{P_{Y_k|X}(y_k|x_j)}{\sum_{j'} p_{j'} P_{Y_k|X}(y_k|x_{j'})} dy_k. \tag{14}$$

For the time sharing scheme using standard constellation, the achievable rate pair $(R_1 + R_2, R_2)$ is such that [1]

$$\begin{cases} R_1 = \alpha \overline{R_1} \\ R_2 = (1-\alpha) \overline{R_2} \end{cases}, \tag{15}$$

where $\overline{R_1}$ and $\overline{R_2}$ are achievable rates for PtP channel using standard M-PAM constellation at SNR_1 and SNR_2, respectively. Varying α from 0 to 1 yields achievable rate region.

Problem (9) is not convex; therefore, direct numerical optimization is inefficient. Clearly, an exhaustive search is not feasible as the complexity would be exponential in the total number of variables. An iterative method for solving (9) is proposed in the next section.

4.2 Numerical solution

Consider a regularized version of (9) as

$$L(P_{UX}, x_0, .., x_{M-1}, s) = \theta \cdot I(X;Y_1|U) + (1-\theta) \cdot I(U;Y_2) + s \cdot (P - \sum_{i=0}^{|\mathcal{U}|-1} \sum_{j=0}^{M-1} p_{ij} \cdot x_j^2), \tag{16}$$

where s is a regularization parameter. For a given value of s, the optimization problem in (16) is solved (for the most general case) with respect to P_{UX} and to $\mathcal{X} = (x_0, x_1, \ldots, x_{M-1})$ alternately until convergence:

$$P_{UX}^{(\ell)} = \arg \max_{P_{UX} \in \mathcal{C}} L(P_{UX}, x_0^{(\ell-1)}, .., x_{M-1}^{(\ell-1)}, s) \tag{17}$$

$$\mathcal{X}^{(\ell)} = \arg\max_{\mathcal{X}} L(P_{UX}^{(\ell)}, x_0, .., x_{M-1}, s), \tag{18}$$

where ℓ is the iteration index, and $\mathcal{C}$ denotes the set of constraints on P_{UX} and can be defined either as $\mathcal{C} = \{P_{UX} : p_{ij} \geq 0, \sum_{i,j} p_{i,j}=1\}$ or as $\mathcal{C} = \{P_{UX}: p_{ij} \geq 0, \sum_i p_{i,j} = \frac{1}{M}\}$ (equiprobable symbols). The optimization problem in (17) with constraint set $\mathcal{C} = \{P_{UX} : p_{ij} \geq 0, \sum_{i,j} p_{i,j} = 1\}$ can be handled by a modified 'Blahut-Arimoto'-type algorithm [29]. Indeed, in order to take into account the regularization, we can show that the Blahut-Arimoto-type algorithm proposed in [30] for broadcast channels should be modified by replacing Equation (19) of Lemma 3 in [30] by $q^*(u,x) = \frac{\beta[Q,\tilde{Q},\bar{Q}](u,x)\cdot e^{-s\frac{x^2}{1-\theta}}}{\sum_{u',x'} \beta[Q,\tilde{Q},\bar{Q}](u',x')\cdot e^{-s\frac{x'^2}{1-\theta}}}$ instead of $q^*(u,x) = \frac{\beta[Q,\tilde{Q},\bar{Q}](u,x)}{\sum_{u',x'} \beta[Q,\tilde{Q},\bar{Q}](u',x')}$, where $\beta[Q,\tilde{Q},\bar{Q}](u,x)$ is defined in Equation (19) of [30]. When there is an additional constraint on constellation symbols to be equiprobable, i.e., $\mathcal{C} = \{P_{UX} : p_{ij} \geq 0, \sum_{i,j} p_{i,j} = 1 \text{ and } \sum_i p_{i,j} = \frac{1}{M}\}$, the Blahut-Arimoto-type algorithm in [30] should also be modified to take into account the additional constraint. In this case, Equation (19) of Lemma 3 in reference [30] should be replaced by $q^*(u,x) = \frac{1}{|\mathcal{X}|}\cdot\frac{\beta[Q,\tilde{Q},\bar{Q}](u,x)}{\sum_u \beta[Q,\tilde{Q},\bar{Q}](u,x)}$, which does not depend on s, where $\beta[Q,\tilde{Q},\bar{Q}](u,x)$ is defined in Equation (19) in this reference.

Now consider (18). The function $L(P_{UX}^{(\ell)}, x_0, .., x_{M-1}, s)$ is not a concave function for all $\mathcal{X} \in \mathbb{R}^M$. However, we observed in our experiments that $L(P_{UX}^{(\ell)}, x_0, .., x_{M-1}, s)$ is a concave function if $\mathcal{X} \in \mathcal{D}$, where $\mathcal{D} = \{\mathcal{X} \in \mathbb{R}^M : |x_i - x_j| > d \ \ \forall i,j \in \{0,..,M-1\} \text{ and } i \neq j\}$, and d depends on the size of the constellation and on the SNR. Since we are interested in finding non-degenerated constellation, we restrict the optimization process to $\mathcal{D}$. Then, a simplex method is used to perform the optimization with initial value in $\mathcal{D}$.

The alternative maximization method can at least increase the objective function in each iteration. In the experiments, we have observed that this method converges at least to a local maximum (denoted $p_{i,j}^*(s), x_j^*(s), 0 \leq j \leq M-1, 0 \leq i \leq |\mathcal{U}|-1$). We discuss now the choice of s. Since we do not know *a priori* which value of s may correspond to the satisfaction of the equality power constraint, we propose to use an iterative process as follows:

$$s^{(k+1)} = \left[s^{(k)} - \gamma \cdot \left(P - \sum_{i=0}^{|\mathcal{U}|-1} \sum_{j=0}^{M-1} p_{ij}^*\left(s^{(k)}\right) \cdot \left(x_j^*\left(s^{(k)}\right)\right)^2 \right) \right]^+, \tag{19}$$

where $[.]^+$ is defined as $[.]^+ = \max(.,0)$. The value of s is increased or decreased with the sign of $P - \sum_{i=0}^{|\mathcal{U}|-1}\sum_{j=0}^{M-1} p_{ij}^*(s^{(k)}) \cdot (x_j^*(s^{(k)}))^2$. The process stops when the power constraint is fulfilled. The proposed algorithm is summarized in Table 2. Obviously, when constellation symbols are constrained to the values of a standard constellation, ($P2$) which is defined in Table 2 will not be used. Similarly, when P_{UX} is uniform, ($P1$) is not used. An alternative interpretation of this algorithm is to recognize that $L(P_{UX}, x_0, .., x_{M-1}, s)$ is the Lagrangian dual of problem 9. Equations (17) and (18) are an iterative method for solving

$$f(s) = \max_{P_{UX}, x_0, .., x_{M-1}} L(P_{UX}, x_0, .., x_{M-1}, s). \tag{20}$$

The dual optimization problem $\min_{\text{s.t. } s\geq 0} f(s)$ is solved in (19) with a gradient-type algorithm. Since $f(s)$ is convex [31], a gradient search method is guaranteed to converge to a global optimum.

5 Result analysis

5.1 Point to point channel

We present in this section the results of maximizing achievable rates for PtP case using M-PAM constellations with $M = 4, 8, 16$ and for different values of SNR. To evaluate the contribution of constellation shaping, we compare, for a fixed SNR, the maximal achievable rate calculated by the algorithm proposed in the previous section to the 'standard constellation' rate, whose symbols are used with equal probability, at the same SNR in terms of SNR saving (called SNR shaping gain). The SNR shaping gain depicted in (Figure 4) is the gain obtained with a fully optimized constellation ($P_{\mathcal{X}}$ and $\mathcal{X}$) compared to the standard M-PAM constellation and when symbols are used with the same probability. To avoid the complexity of constructing nearly optimal input distribution codes, another method for doing constellation shaping is to optimize only the position of symbols in the constellation. Each signal point is assumed to be chosen with the same probability; however, the position of each point in the constellation is optimized. The corresponding shaping gain is given in (Figure 5). We observe the following: the shaping

Table 2 Numerical solution for solving (9)

Step		Solution
Step 0		$s \leftarrow s^{(0)}$
Step k	Step 0	$\mathcal{X} \leftarrow \mathcal{X}^{(0)}$ where $\mathcal{X} = (x_0, x_1, .., x_{M-1})$
	Step ℓ	$P_{UX}^{(\ell)} = \arg\max_{P_{UX}\in\mathcal{C}} L(P_{UX}, \mathcal{X}^{(\ell-1)}, s^{(k-1)})$ ($P1$) $\mathcal{X}^{(\ell)} = \arg\max_{\mathcal{X}} L(P_{UX}^{(\ell)}, \mathcal{X}, s^{(k-1)})$ ($P2$)
	Stopping criterion	$\lvert L(P_{UX}^{(\ell)}, \mathcal{X}^{(\ell)}, s^{(k)}) - L(P_{UX}^{(\ell-1)}, \mathcal{X}^{(\ell-1)}, s^{(k-1)})\rvert \leq \epsilon_L$
		$s^{(k)} = [s^{(k-1)} - \beta(P - \sum_{i,j} p_{ij}^*(s^{(k-1)}) \cdot (x_j^*(s^{(k-1)}))^2)]^+$ where $[.]^+ = \max(.,0)$
Stopping criterion		$\lvert s^{(k)} - s^{(k-1)}\rvert \leq \epsilon_s$

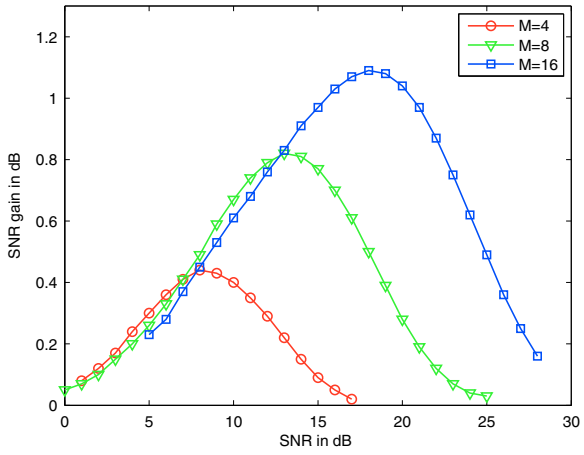


Figure 4 SNR shaping gain in dB for PtP channel. $\mathcal{X}$ and $P_{\mathcal{X}}$ are optimized.

gain depends on the SNR and on the size of the constellation. The maximum gain is obtained for mid-range SNR. The distribution of probability $P_{\mathcal{X}}$ (not reported) is very similar to the sampling of a Gaussian distribution. With the half-optimized constellation ($\mathcal{X}$ only), a significant degradation is observed for mid-range SNR compared to that for the fully optimized constellation. Hence, we can conclude that symbol pdf optimization is useless at low and high SNR, whereas the fully optimized constellation is efficient for mid-range SNR, in which case the gain increases with the size of the constellation.

5.2 Broadcast channel

Current broadcast systems are using two practical transmission schemes for sending information to users: orthogonal schemes in which the time and/or frequency is split between the users, and superposition modulation schemes where the constellation for each user is fixed. In this section, a comparison is provided between these standard schemes and various (more complex) transmission strategies such as superposition coding. The effect of constellation shaping is evaluated by analyzing the achievable rate region curves obtained for an M-PAM constellation ($M = 4, 8, 16$) and for several pairs (SNR_1, SNR_2). The following schemes are considered:

- Time sharing using standard M-PAM (TS).
- SM - 3 possible configurations (see Table 1)
- SC - 4 possible configurations (see Table 1)

In the following, we denote by the 'case 1' of superposition modulation when $M_1 = 2, M_2 = 4$ and when $M_1 = 2, M_2 = 8$. 'Case 2' is when $M_1 = 4, M_2 = 2$ and when $M_1 = 4, M_2 = 4$. 'Case 3' refers to the case when $M_1 = 8$, $M_2 = 2$.

Achievable rate region curves are provided in Figures 6, 7, 8, 9, 10, and 11 for $M = 4, 8, 16$. For each value of M, the display of the results is limited to two different pairs of SNR. In complement with the achievable rate region curves, comparisons are also conducted in terms of SNR savings for target achievable rates (maximum shaping gain) and in terms of maximum percentage of gain for user 1. These two quantities are defined below.

Definition 1. Consider two transmission strategies (A and B). The pair of rates $(R_1 + R_2, R_2)$ is achieved for $(\mathrm{SNR}_1, \mathrm{SNR}_2)$ with A and for $(\mathrm{SNR}_1 + \Delta\mathrm{SNR}, \mathrm{SNR}_2 + \Delta\mathrm{SNR})$ with B. The shaping gain (with A compared to B) is ΔSNR. The maximum shaping gain is defined as

$$\mathrm{MG}_{\mathrm{SNR}_{dB}}(A|B) = max_{R_2} \Delta\mathrm{SNR} \tag{21}$$

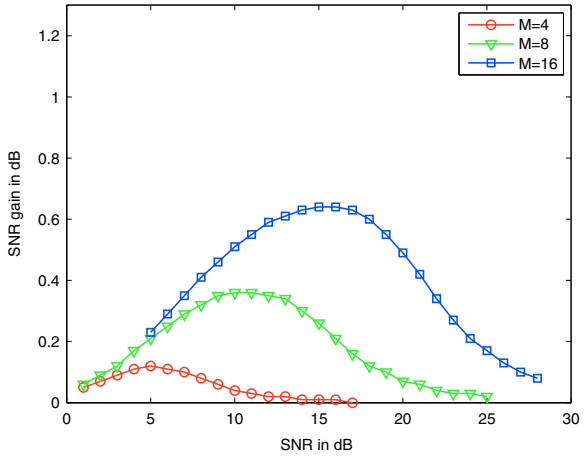


Figure 5 SNR shaping gain in dB for PtP channel where only $\mathcal{X}$ is optimized.

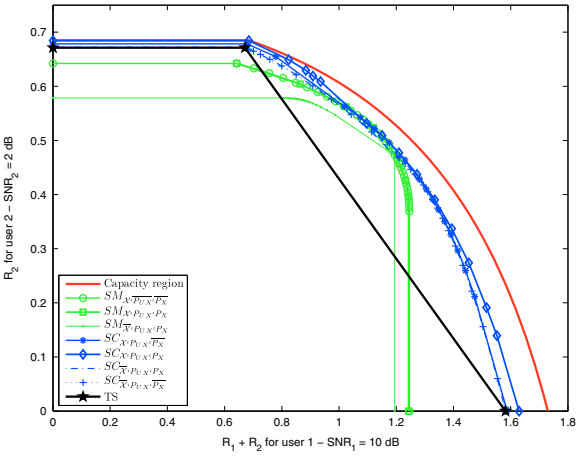


Figure 6 Achievable rate regions with $M = 4$ and $(\mathrm{SNR}_1, \mathrm{SNR}_2) = (10\ \mathrm{dB}, 2\ \mathrm{dB})$.

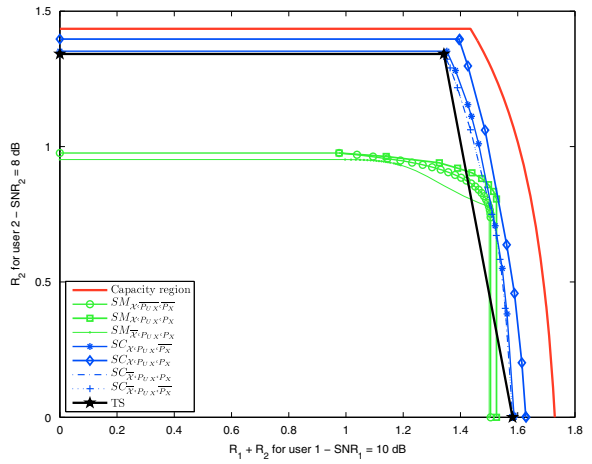


Figure 7 Achievable rate regions with $M = 4$ and $(\text{SNR}_1, \text{SNR}_2) = (10\text{ dB}, 8\text{ dB})$.

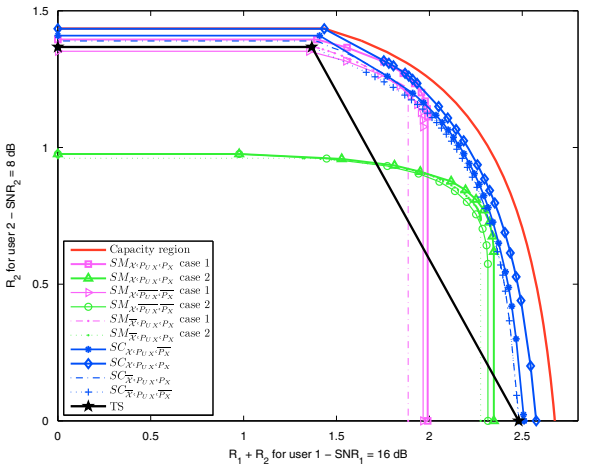


Figure 9 Achievable rate regions with $M = 8$ and $(\text{SNR}_1, \text{SNR}_2) = (16\text{ dB}, 8\text{ dB})$.

Definition 2. Consider two transmission strategies (A and B). For a given pair of SNR $(\text{SNR}_1, \text{SNR}_2)$ and a fixed value of R_2, the achievable pair of rates is $(R_1^A + R_2, R_2)$ and $(R_1^B + R_2, R_2)$ with A and B, respectively. The gain on the achievable rate for user 1 is given by

$$G_{R_1}(A|B) = \frac{\left(R_1^A + R_2\right) - \left(R_1^B + R_2\right)}{R_1^B + R_2} \cdot 100\ (\%). \qquad (22)$$

The maximum gain on the achievable rate for user 1 (with A compared to B) is given by

$$\text{MG}_{R_1}(A|B) = max_{R_2} G_{R_1}(A, B). \qquad (23)$$

5.2.1 Superposition modulation

In this section, the three possible configurations of superposition modulation are compared. We can see from Figures 6, 7, 8, 9, 10, and 11 that $\text{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$ (optimization of $\mathcal{X}$ only) outperforms $\text{SM}_{\overline{\mathcal{X}},P_{UX},P_X}$ (optimization of P_{UX} only) in terms of maximal achievable rates per user when $M = 4$. For $M = 8$ and 16, $\text{SM}_{\overline{\mathcal{X}},P_{UX},P_X}$ can achieve slightly higher rates than $\text{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$. The implementation of a system with constellation symbols with non-standard positions and generated with the same probability is less complex than the implementation of a system which generates symbols with non-uniform joint distribution of probability. Thus, $\text{SM}_{\overline{\mathcal{X}},P_{UX},P_X}$ does not seem to be of interest since it is not very efficient in terms of achievable rates and is more complex to implement.

Figures of achievable rate region show that an improvement can be obtained with $\text{SM}_{\mathcal{X},P_{UX},P_X}$ (full optimization) compared to $\text{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$ (optimization of $\mathcal{X}$ only) and depending on $\delta_{\text{SNR}} = \text{SNR}_1 - \text{SNR}_2$. Numerical values of

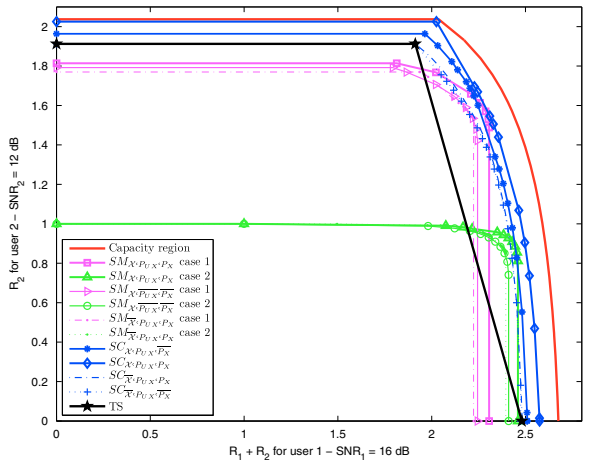


Figure 8 Achievable rate regions with $M = 8$ and $(\text{SNR}_1, \text{SNR}_2) = (16\text{ dB}, 12\text{ dB})$.

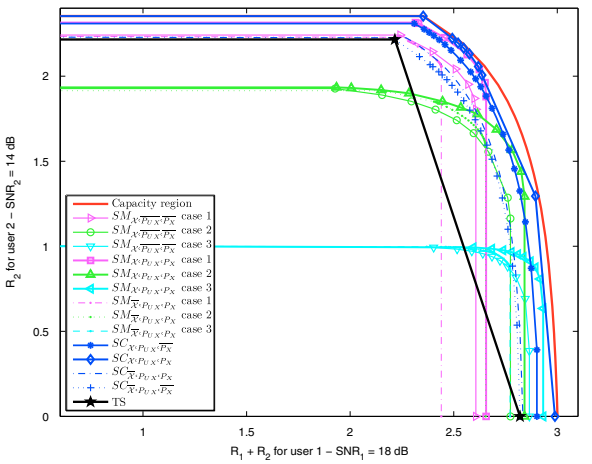


Figure 10 Achievable rate regions with $M = 16$ and $(\text{SNR}_1, \text{SNR}_2) = (18\text{ dB}, 14\text{ dB})$.

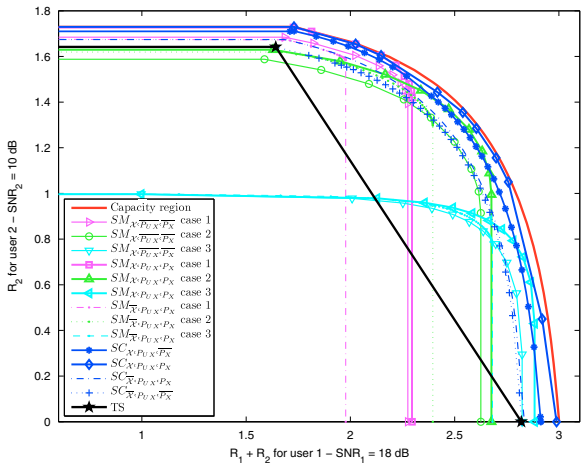


Figure 11 Achievable rate regions with $M = 16$ and $(\mathrm{SNR}_1, \mathrm{SNR}_2) = (18\ \mathrm{dB}, 10\ \mathrm{dB})$.

the maximum gain in achievable rate (MG_{R_1}) and of the maximum SNR savings ($\mathrm{MG}_{\mathrm{SNR}_{dB}}$) are given in Table 3. We observe the following: a slight gain in terms of achievable rates can be translated into a noticeable gain in terms of SNR saving. The maximum shaping gain increases with the constellation size. Thus, constellation shaping for the SM strategy seems more useful for high values of M. The analysis of the optimal matrix P_{UX} (results not reported) leads to the conclusion that X_1 and X_2 are not independent in general when using finite-size constellations. We observe also that the maximum shaping gain for $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ versus $\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$ increases when δ_{SNR} decreases, independently of M. In particular, full optimization (vs. optimization of the symbol position) does not provide significant improvement for large SNR gap in the SM strategy.

Table 3 Comparison of $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ (A) and $\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$ (B) with respect to $\mathrm{MG}_{\mathrm{SNR}_{dB}}$ and MG_{R_1}

M	SNR_1	SNR_2	$\mathrm{MG}_{\mathrm{SNR}_{dB}}(A\|B)$	$\mathrm{MG}_{R_1}(A\|B)$
		8	0.39	7.46%
4	10	6	0.17	3.51%
		4	0.05	1.77%
		2	0.01	0.38%
		14	$0.71^{(M_1=4,M_2=2)}$	$20.17\%^{(M_1=4,M_2=2)}$
8	16	12	$0.57^{(M_1=4,M_2=2)}$	$13.21\%^{(M_1=4,M_2=2)}$
		10	$0.41^{(M_1=4,M_2=2)}$	$13.07\%^{(M_1=2,M_2=4)}$
		8	$0.33^{(M_1=2,M_2=4)}$	$18.93\%^{(M_1=2,M_2=4)}$
		16	$1.05^{(M_1=8,M_2=2)}$	$10.67\%^{(M_1=8,M_2=2)}$
16	18	14	$0.87^{(M_1=8,M_2=2)}$	$11.54\%^{(M_1=8,M_2=2)}$
		12	$0.64^{(M_1=8,M_2=2)}$	$12.08\%^{(M_1=4,M_2=4)}$
		10	$0.49^{(M_1=8,M_2=2)}$	$19.53\%^{(M_1=4,M_2=4)}$

5.2.2 Time sharing or superposition modulation?

This section compares two strategies (TS and SM) classically considered in broadcast systems. In Figures 6 and 7 ($M = 4$), we observe that the achievable rate region can be split into two parts. Indeed, for small and large values of R_2, TS is better than SM. On the contrary, SM is better than TS for middle-range values of R_2. Under a given rate requirement for one user, we can thus determine the best transmission strategy. We can also observe that the region in which SM is better than TS becomes small for larger values of SNR_2. With $M = 8$ (Figures 8 and 9), the area in which SM is better than TS increases (compared to $M = 4$) by considering the union of the two possible configurations for SM: $M_1 = 2, M_2 = 4$ (case 1) and $M_1 = 4, M_2 = 2$ (case 2). This is particularly true when δ_{SNR} increases. We also observe that TS can achieve higher rates than SM (case 1) for good SNR_2 values. Indeed, the maximum rate of user 2 with SM is the maximum individual rate for a 4-PAM constellation, whereas it is the individual user rate that achieved using standard 8-PAM in the TS case. For low SNR_2 values, optimized 4-PAM may achieve higher rate than standard 8-PAM; thus, SM becomes better in this interval. For a 16-PAM constellation (Figures 10 and 11), SM is always better than TS for the studied pairs of $(\mathrm{SNR}_1, \mathrm{SNR}_2)$. Table 4 shows the maximum percentage of improvement in achievable rate of user 1 by TS when using $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ (full optimization) strategy in the interval, where $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ is better than TS. Clearly, the maximum percentage of improvement increases when δ_{SNR} increases, and an important gain is obtained for high values of δ_{SNR} as in the case of $\mathrm{SNR}_1 = \delta_{\mathrm{SNR}} = 10$ dB for a 4-PAM, where the percentage of gain on achievable rate of user 1 varies between 0% and 40.7%. For a 8-PAM constellation, the percentage of gain

Table 4 Comparison of $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ (A) vs. TS (B) and comparison of $\mathrm{SC}_{\mathcal{X},P_{UX},P_X}$ (A) vs. TS $\bigcup$ $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ (C)

M	SNR_1	SNR_2	$\mathrm{MG}_{R_1}(A\|B)$	$\mathrm{MG}_{R_1}(A\|C)$
		8	6.13%	6.72%
4		6	11.14%	11.65%
	10	4	18.50%	16.69%
		2	28.43%	18.9%
		0	40.70%	23.54%
		14	$7.80\%^{(M_1=2,M_2=4)}$	7.89%
8	16	12	$13.60\%^{(M_1=2,M_2=4)}$	11.43%
		10	$21.15\%^{(M_1=2,M_2=4)}$	14.96%
		8	$30.21\%^{(M_1=2,M_2=4)}$	14.71%
		16	$10.36\%^{(M_1=2,M_2=8)}$	2.96%
16	18	14	$16.42\%^{(M_1=4,M_2=4)}$	2.94%
		12	$24.68\%^{(M_1=4,M_2=4)}$	5.29%
		10	$35.08\%^{(M_1=4,M_2=4)}$	4.80%

on achievable rate of user 1 varies between 0% and 30.21% when $\mathrm{SNR}_1 = 16$ dB and $\delta_{\mathrm{SNR}} = 8$ dB. For a 16-PAM, the percentages of improvements can be up to 35.08% when $\mathrm{SNR}_1 = 18$ dB and $\delta_{\mathrm{SNR}} = 8$ dB. We can conclude that SM is a better option than TS especially for large δ_{SNR} values. TS is optimal in the region, where we want to maximize the rate of user 2 for good values of SNR_2 because the single user rate achieved by TS is the rate achieved using standard M-PAM constellation (the constellation is split between users with SM). Thus, SM seems more gainful than TS when we want to serve users with very diverse SNRs.

5.2.3 Is superposition coding necessary?

For the three constellations under consideration ($M = 4, 8, 16$), the maximal achievable rate region obtained by the optimal general case of superposition coding when we consider the general form of P_{UX} (SC) can achieve, depending on M and user SNRs, a large region of rate pairs $(R_1 + R_2, R_2)$ that cannot be achieved neither by TS nor by SM. Even when we fully optimize SM ($\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$), we are far from maximal achievable rate region. Sometimes, the maximal achievable rate region curve is very close or even coincides with the $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ achievable rate region in a pair of rates $(R_1^* + R_2^*, R_2^*)$. This is the case when $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ is the optimal superposition coding in terms of achievable rates. We can see for example in Figure 6 that the pair of rates $(R_1^* + R_2^* = 1.096, R_2^* = 0.531$ which corresponds to the optimal rate pair when we optimize the general case of SC for $\theta = 0.23$) is an intersection point with $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ achievable rate region.

We are interested now in the numerical evaluation of the gain in rate of user 1 ($R_1 + R_2$) when we use $\mathrm{SC}_{\mathcal{X},P_{UX},P_X}$ (full optimization) compared to the best strategy between TS and SM. This gain ($\mathrm{MG}_{R_1}(\mathrm{SC}_{\mathcal{X},P_{UX},P_X} | TS \bigcup \mathrm{SM}_{\mathcal{X},P_{UX},P_X})$ calculated in % is the distance between the limit of the maximal achievable rate region and the limit of the union of achievable rate regions of TS and $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$.

The results are reported in Table 4. We observe that the part of the maximal achievable rate region which is unachievable by TS and SM is bigger when M is small because we observe that for the case of 4-PAM, we have one configuration for SM. However, we have two configurations of SM for 8-PAM constellation and three configurations for 16-PAM constellation. Thus, when M increases, the union of achievable rates for all SM cases tends to the sets of achievable rates by the general superposition coding. Asymptotically, we know that when $M \to \infty, \mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ is the optimal superposition coding scheme because it allows the capacity region for two-user AWGN BC using Gaussian alphabet for each user to be achieved. Thus, the maximum gain in user 1 rate decreases when the constellation order M increases. We observe also that the gain in achievable rates is high for high values of δ_{SNR}. On the other hand, the experiments show that by using the general superposition coding strategy with the constraint that symbols should be equiprobable ($\mathrm{SC}_{\mathcal{X},P_{UX},\overline{P_X}}$), the loss is limited compared to the full optimization ($\mathrm{SC}_{\mathcal{X},P_{UX},P_X}$), 4.84%, 7.66%, and 3.94% for the simulated pairs of $(\mathrm{SNR}_1, \mathrm{SNR}_2)$ when $M = 4$, 8, and 16, respectively. This means that we can use equiprobable symbols with, in general, a small loss in achievable rates. However, $\mathrm{SC}_{\mathcal{X},P_{UX},\overline{P_X}}$ is not an interesting case when $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$ can achieve better rates since SM is less complex to implement than SC.

Moreover, with standard M-PAM symbols, the two possible configurations ($\mathrm{SC}_{\overline{\mathcal{X}},P_{UX},P_X}$ (optimization of P_{UX} and P_X) and $\mathrm{SC}_{\overline{\mathcal{X}},P_{UX},\overline{P_X}}$ (optimization of P_{UX} only)) give very similar results in most considered pairs of SNR. We also observe that the loss in maximum achievable rate experienced by user 1 with $\mathrm{SC}_{\overline{\mathcal{X}},P_{UX},P_X}$ is less than 10% under the rate experienced with $\mathrm{SC}_{\mathcal{X},P_{UX},P_X}$. Thus, we can use standard values of symbol positions without losing much on achievable rates.

In general, one can conclude that fixing constellations of users (i.e., assigning labels to the constellation so that we distinguish between the bits intended for each user) is not optimal for coding and may result in important loss in terms of rates for systems using finite-size constellations especially for low-order constellations. A better solution is to determine the optimal alphabet of the auxiliary alphabet U which is not necessarily a constellation and then to generate the codewords x^n which are not necessarily the sum of two codewords (see Section 3.4).

6 Application: coverage extension

We first consider a transmission over a broadcast channel with finite size input alphabet. For simplicity of the illustration and without loss of generality, let us assume that the existing user alphabet belongs initially to a standard constellation whose symbols are used with equal probability. We assume that the existing user is at distance d_0 from the sender achieving a rate R_0. Some information is also to be transmitted to an upgraded layer of users. The sender can use up to 16 symbols, then several transmission schemes can be used. We are interested in comparing the transmission schemes to serve the new user under two scenarios: either the new user is closer to the transmitter than the existing user or the new user is farther than the existing one. For a target rate R_0 that is fixed for the existing user and achievable using a standard M-PAM and equiprobable symbols, we are interested in determining the variation of the coverage's diameter ratio between the two layer of users as a

function of the achievable rate by the upgraded user for various broadcast transmission strategies. We assume that SNR $\propto \frac{1}{d^2}$.

6.1 The sender can use up to 16 symbols

6.1.1 Scenario 1

In this scenario, the system consists initially of one layer of users. Now, assume that the data information is also to be transmitted to a second layer of users with higher SNR. In the following, we keep the notation from the preceding section, where the user with greater SNR is denoted by user 1. Thus, in this scenario, the legacy receivers are denoted by user 2 which is at a distance d_2 from the transmitter and achieving a rate R_0 when the data is modulated using standard 4-PAM constellation and equiprobable symbols. The upgraded receivers are denoted by user 1 ($\text{SNR}_1 > \text{SNR}_2$). We intend that the good user receives more throughput than user 2 via the use of 16-PAM.

In this example SNR_2 is fixed to 10 dB. Initially, user 2's alphabet belongs to a 4-PAM standard constellation (see section 3.1), and the rate transmitted to user 2 is $R_0 = 1.582$ bits/ch. use.

Now, a new layer of users called user 1 is introduced in the system with $\text{SNR}_1 > \text{SNR}_2$. Our target is to provide the maximum bit rate to the new user without changing R_0 or d_0 and using a 16-PAM. By enlarging the constellation and optimizing the symbol positions and probability distribution, we ensure that the rate of the initial user will not decrease after introducing a new user.

Consider now the results for the following strategies which can achieve a positive private-message rate for user 1: time sharing using standard 16-PAM, $\text{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}} M_2 = 8/M_1 = 2$ (optimization of $\mathcal{X}$ only), $\text{SM}_{\mathcal{X},P_{UX},P_X} M_2 = 8/M_1 = 2$ (full optimization) and $\text{SC}_{\mathcal{X},P_{UX},P_X}$ (full optimization). Figure 12 illustrates the variation of d_1/d_2, which is the ratio of the diameter of the coverage area for user 1 over the diameter of the initial coverage area for user 2, as a function of the achievable rate for user 1 for a target rate $R_0 = 1.582$ for user 2.

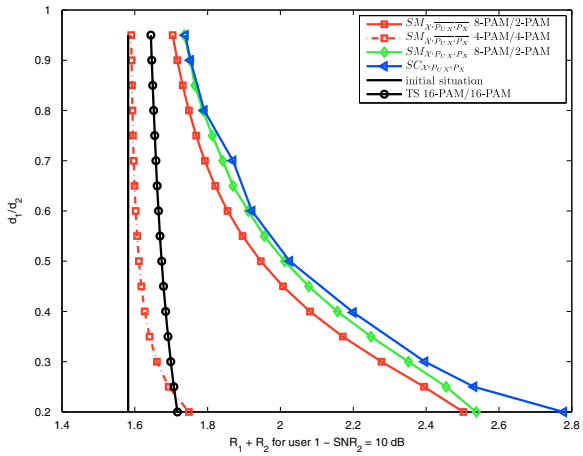


Figure 12 Coverage ratio d_1/d_2 as function of the achievable rate for user 1.

Let us assume for example that the new user is midway between the transmitter and user 2 ($d_1/d_2 = 0.5$). Figure 12 shows that the most simple case of superposition modulation ($\text{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}} M_2 = 8/M_1 = 2$) provides 16.3% more bit rate than time sharing for the new user. If we move immediately to a more complex case and optimize P_{UX} ($\text{SM}_{\mathcal{X},P_{UX},P_X} M_2 = 8/M_1 = 2$), a gain of 21% is obtained on the bit rate of user 1 compared to time sharing. This gain on achievable rate for the new user is equivalent to a gain of 1 dB on SNR_1 compared to superposition modulation with uniform P_{UX}. However, if we move to the most general case of superposition coding, it does not provide significant gain compared to superposition modulation.

Now, we assume that the new user is close to the transmitter such that $d_1/d_2 = 0.2$. We observe that the gain on the bit rate of user 1 using the simple case of superposition modulation increases to 45.7% compared to time sharing. By moving to a more complex case ($\text{SM}_{\mathcal{X},P_{UX},P_X} M_2 = 8/M_1 = 2$), a gain of 47.8% is obtained on the bit rate of user 1 compared to time sharing. We observe also that it is relevant in this case to move to the most general case of superposition coding since it provides a gain of 61.8% on the bit rate of user 1 compared to time sharing.

Consequently, using superposition modulation provides always noticeable gain compared to time sharing. The general case of superposition coding $\text{SC}_{\mathcal{X},P_{UX},P_X}$ is useful when user 1 is close to the transmitter, but not when it is close to user 2.

6.1.2 Scenario 2

Initially, consider a system of one layer of users, denoted by user 1, at a distance d_1 from the transmitter and achieving a rate R_0. Moreover, the alphabet of user 1 belongs to a standard 8-PAM constellation. In this example, SNR_1 is fixed to 18 dB. Thus, user 1 can achieve a rate $R_0 = 2.73$ bits/ch. use in the initial situation. In this scenario, we want to serve a second layer of users denoted by user 2 which is farther to the transmitter than the existing user, i.e., $\text{SNR}_2 < \text{SNR}_1$.

Achievable rates for user 2 are obtained at different distance d_2 from the transmitter and using various transmission strategies for a target rate of user 1 equal to R_0 and a coverage diameter for user 1 fixed to d_1. Figure 13 illustrates the variation of d_2/d_1, which is the ratio of the diameter of the coverage area for user 2 over the diameter of the initial coverage area for user 1, as a function of the achievable rate for user 2 when a target rate for user 1 is fixed to $R_0 = 2.73$ bits/ch. use.

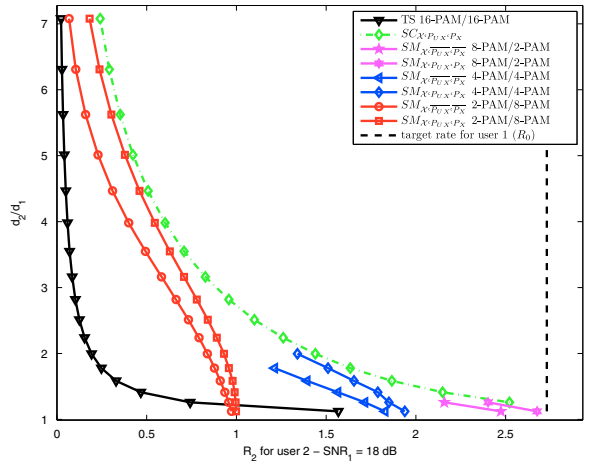


Figure 13 Coverage ratio d_2/d_1 as function of the achievable rate for user 2.

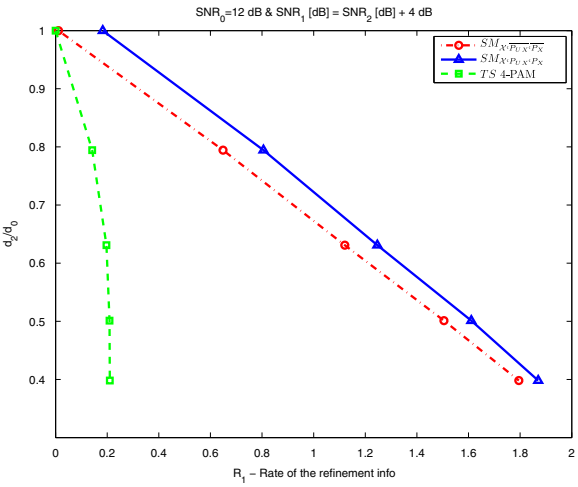


Figure 14 Reduction in legacy coverage d_2/d_0 in function of the rate of the refinement R_1.

We observe in Figure 13 that superposition modulation can always achieve better rates for user 2 than time sharing using 16-PAM. Let us assume first that we want to increase the diameter of the coverage area for the new user (user 2) such that $d_2/d_1 = 4$. Time sharing provides a bit rate less than 0.06 bits/ch. use. The most simple case of superposition modulation ($\mathrm{SM}_{\mathcal{X},P_{UX},P_X} M_2 = 2/M_1 = 8$) provides a significant improvement on the achievable rate for user 2 which is equal to 0.4 bits/ch. use in this case. If we increase the complexity by optimizing the joint probability distribution P_{UX}, we obtain 35% more bit rate for user 2 comparing to superposition modulation with uniform P_{UX}. If we move to the general case of superposition coding, we gain only 10% on the bit rate of the new user compared to superposition modulation (see Table 5). However, when the new layer of users is at distance $d_2 = 2.25\ d_1$, the general case of superposition coding provides a significant gain of 41% on the achievable rate of user 2 comparing to superposition modulation.

Consequently, the general case of superposition coding can bring significant gains compared to superposition modulation, depending on the diameter of the coverage area for the new layer of users. For superposition modulation, optimizing the joint distribution of probability P_{UX} provides often significant shaping gains.

Table 5 Comparison of $\mathrm{SC}_{\mathcal{X},P_{UX},P_X}$ and $\mathrm{SM}_{\mathcal{X},P_{UX},P_X} M_2 - \mathrm{PAM}/M_1 - \mathrm{PAM}$ w.r.t MG_{R_2} (%)

d_2/d_1	SNR_2	MG_{R_2}	M_2/M_1
1.2589	16	4.9416	8/2
1.4125	15	20.1521	4/4
1.5849	14	12.7522	4/4
1.7783	13	8.2192	4/4
1.9953	12	7.4536	4/4
2.2387	11	41.4993	2/8
2.5119	10	30.8293	2/8
2.8184	9	22.9121	2/8
3.1623	8	16.7443	2/8
3.5481	7	12.6033	2/8
3.9811	6	10.5427	2/8
4.4668	5	10.3343	2/8
5.0119	4	11.7414	2/8
5.6234	3	16.0961	2/8
6.3096	2	22.8535	2/8
7.0795	1	32.6194	2/8

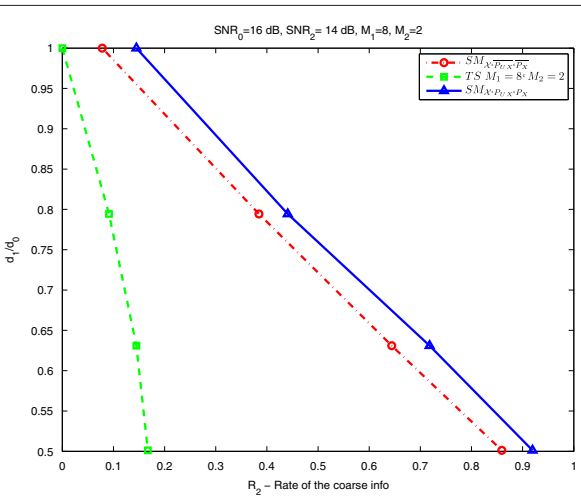


Figure 15 Reduction in legacy coverage d_1/d_0 in function of the rate of the coarse information R_2.

6.2 The cardinality of the existing user alphabet is kept fixed :

In this section, we study scenario 1 (and 2) supposing that the legacy receivers will continue working as in the initial situation, still using 4-PAM (8-PAM). The system consists initially one layer of users at distance d_0 from the transmitter and achieves a rate R_0. Now, we want to change the transmitter, such that the upgraded receivers closer (farther) in range will be able to decode a refinement (coarse) layer and use a 16-PAM constellation. Thus, only time sharing with $M_1 = M_2 = 4 (M_1 = 8, M_2 = 2)$ and superposition modulation strategies can be used. We aim to study how small the reduction in legacy coverage can be made, depending on the rate of the refinement (coarse) information achieved by the upgraded users. Thus, suppose that the legacy coverage can be reduced from d_0 to d_2 (from d_0 to d_1). We have studied this problem for $\mathrm{SNR}_0 = 12$ dB and for $\mathrm{SNR}_1 - \mathrm{SNR}_2 = 4$ dB in scenario 1 (and for $\mathrm{SNR}_0 = 16$ and $\mathrm{SNR}_2 = 14$ dB in scenario 2). Figures 14 and 15 represent the reduction in coverage d_2/d_0 (and d_1/d_0 respectively) as a function of the rate of the refinement R_1 (of the coarse R_2), while the rate achieved by the legacy receivers is kept fixed to its initial situation, i.e., R_0.

We observe in Figures 14 and 15 that the gain of superposition modulation strategies over time sharing becomes more important when d_2/d_0 (d_1/d_0) is small. These figures show that using superposition modulation when both symbol positions and P_{UX} are optimized, we gain around 5% from the initial coverage compared to the case of superposition modulation where symbols are used with equal probability. We can observe also that a reduction of only 10% and 20% in coverage area for the existing user can serve the upgraded user with a rate up to 20% and 35% (9% and 15%) from the rate achieved by the legacy users, using $\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$. Consequently, by using $\mathrm{SM}_{\mathcal{X},\overline{P_{UX}},\overline{P_X}}$, the legacy receivers still use 4-PAM (8-PAM in scenario 2), and we can serve a new layer of users with an acceptable rate, a small reduction in coverage area, and with less complexity compared to $\mathrm{SM}_{\mathcal{X},P_{UX},P_X}$.

7 Conclusion

In this work we considered the problem of maximizing the achievable rate region for power-constrained AWGN broadcast channel of two users using M-PAM constellations. The achievable rate region is given for various transmission strategies. Maximal achievable rate region for superposition coding and superposition modulation is obtained using constellation shaping. An iterative algorithm was proposed to solve this optimization problem. Then, the efficiency of several strategies are compared. For superposition modulation, the results showed that constellation shaping seems more useful for high values of M. Moreover, the gain in using a complex case of superposition modulation increases when the SNR gap between users decreases. We observed also that superposition modulation outperforms time sharing in a large part of the achievable rate region. On the other hand, it is shown that using the general case of superposition coding can bring important gains compared to classical schemes. We observed also that in the case of finite input alphabet, superposition modulation is not the optimal strategy as in the case of Gaussian input alphabets. Finally, in order to make clear that this paper provides useful tools for the system designer, we considered two scenarios of coverage areas and user alphabets where the systems served initially one layer of users. Then, we propose to serve a second layer of users, and we evaluate the achievable rate of the new layer depending on the broadcast strategy. To improve the system performance compared to time sharing, we can optimize the joint probability distribution and symbol positions of the superimposed modulations or consider the general case of superposition coding. In this work, we showed that the optimization of probabilities was often useful, but not always. However, superposition coding brings sometimes significant gains compared to superposition modulation, depending on the diameter of coverage area for the new layer of users.

This work can also be extended to two-dimensional constellations like M-QAM and other channel models. The maximization achievable rates using various transmission strategies can be performed also using the proposed algorithm based on alternative maximization with respect to symbol positions and the joint distribution of probability.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]University Paris-Sud, UMR8506 Orsay, F-91405, France. [2]CNRS, Gif-sur-Yvette, F-91192, France. [3]Supelec, Gif-sur-Yvette, F-91192, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette Cedex, France.

Received: 27 June 2013 Accepted: 20 October 2013
Published: 31 October 2013

References

1. TM Cover, Broadcast channels. IEEE Trans. Inform. Theory **18**, 2–14 (1972)
2. PP Bergmans, Random coding theorem for broadcast channels with degraded components. IEEE Trans. Inform. Theory **19**(2), 197–207 (1973)
3. PP Bergmans, A simple converse for broadcast channels with additive white Gaussian noise. IEEE Trans. Inform. Theory **20**, 279–280 (1974)
4. RG Gallager, Capacity and coding for degraded broadcast channels. Probl. Infor. Transm. **10**(3), 185–193 (1974)
5. G Imai, S Hirakawa, A new multilevel coding method using error correcting codes. IEEE Trans. Inform. Theory **23**, 371–377 (1977)
6. G Ungerboeck, Channel coding with multilevel/phase signals. IEEE Trans. Inform. Theory **28**, 55–67 (1982)
7. PP Bergmans, TM Cover, Cooperative broadcasting. IEEE Trans. Inform. Theory **20**, 317–324 (1974)
8. European Telecommunications Standards Institute, *EN 300 744: Digital Video Broadcasting (DVB)—framing structure, channel coding and modulation for digital terrestrial television*. (European Telecommunications Standards Institute, France, 2004–2006)

9. European Telecommunications Standards Institute, *ETSI TS 102: Digital Video Broadcasting (DVB)—system specifications for satellite services to handheld devices (SH) below 3 GHz*. (European Telecommunications Standards Institute, France, 2008), p. 585
10. H Meric, J Lacan, C Amiot-Bazile, F Arnal, ML Boucheret. Generic approach for hierarchical modulation performance analysis: application to DVB-SH, in *Wireless Telecommunications Symposium* (New York, 13–15 April 2011)
11. AR Calderbank, LH Ozarow, Nonequiprobable signaling on the Gaussian channel. IEEE Trans. Inform. Theory **36**(4), 726–740 (1990)
12. D Sommer, G Fettweis. Shaping by non-uniform QAM for AWGN channels and applications using turbo coding, in *ITG Conference Source and Channel Coding* (Munich, Germany, 17–19 Jan 2000)
13. C Fragouli, RD Wesel, D Sommer, GP Fettweis, Turbo codes with non-uniform constellations. Proc. IEEE Int. Conf. Commun. **1**, 70–73 (2001)
14. N Varnica, X Ma, A Kavcic, Capacity of power constrained memoryless AWGN channels with fixed input constellations. GLOBECOM **2**, 1339–1343 (2002)
15. D Raphaeli, A Gurevitz, Constellation shaping for pragmatic turbo-coded modulation with high spectral efficiency. IEEE Trans. Commun. **52**(3), 341–345 (2004)
16. SY LeGoff, BK Khoo, CC Tsimenidis, BS Sharif, Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver. IEEE Trans. Wireless Commun. **6**(6), 2223–2233 (2007)
17. NH Ngo, SA Barbulescu, SS Pietrobon. Performance of nonuniform M-ary QAM constellation on nonlinear channels, in *Australian Communications Theory Workshop* (Australia, 2–4 Feb 2005)
18. J Zhang, D Chen, Y Wang. A new constellation shaping method and its performance evaluation in BICM-ID, in *Vehicular Technology Conference Fall (VTC 2009-Fall)* (Anchorage, AK, 20–23 Sept 2009)
19. M Valenti, X Xiang, Constellation shaping for bit-interleaved LDPC coded APSK. IEEE Trans. Commun. **60**(10), 2960–2970 (2012)
20. C Huppert, M Bossert. On achievable rates in the two user AWGN broadcast channel with finite input alphabets, in *ISIT* (Nice, 24–29 June 2007)
21. TM Cover, JA Thomas, *Elements of Information Theory*, 2nd edn. (Wiley, Hoboken, 2006)
22. J Gledhill, P Macavock, R Miles, *DVB-T: Hierarchical Modulation*. (DVB, Geneva, 2000)
23. A Schertz, C Weck, *Technical Review: Hierarchical Modulation-the Transmission of Two Independent DVB-T Multiplexes on a Single Frequency*. (EBU, Switzerland, 2003)
24. V Singh. On superposition coding for wireless broadcast channels. Master's thesis, Royal Institute of Technology, Sweden (2005). www.ee.kth.se/php/modules/publications/reports/2005/IR-SB-EX-0507.pdf
25. Z Mheich, P Duhamel, L Szczecinski, ML Alberi-Morel. Constellation shaping for broadcast channels in practical situations, in *19th European Signal Processing Conference* (Barcelona, 29 Aug–2 Sept 2011)
26. Z Mheich, ML Alberi-Morel, P Duhamel, Optimization of unicast services transmission for broadcast channels in practical situations. Bell Labs Techn. J. **17**(5–24) (2012)
27. Z Mheich, F Alberge, P Duhamel. On the efficiency of transmission strategies for broadcast channels using finite size constellations, in *21st European Signal Processing Conference* (Marrakech, 9–13 Sept 2013)
28. TM Cover, Comments on broadcast channels. IEEE Trans. Inform. Theory **44**(6), 2524–2530 (1998)
29. RE Blahut, Computation of channel capacity and rate-distortion functions. IEEE Trans. Inform. Theory **18**(4), 460–473 (1972)
30. K Yasui, T Matsushima. Toward computing the capacity region of degraded broadcast channel, in *ISIT* (Austin, TX, 13–18 June 2010)
31. DP Bertsekas, *Nonlinear Programming*, 2nd edn. (Athena Scientific, Nashua, 1999)

doi:10.1186/1687-1499-2013-254
Cite this article as: Mheich *et al.*: **Achievable rates optimization for broadcast channels using finite size constellations under transmission constraints.** *EURASIP Journal on Wireless Communications and Networking* 2013 **2013**:254.

Annexe 4

F. Alberge, Z. Naja, P. Duhamel.
Power extrinsic propagation for turbo-codes. IEEE Trans. on Signal Processing, vol 61, n° 5, 2013.

Abstract

A generalized information passing strategy is discussed for iterative decoding of turbo-like systems. Both the existence of fixed point and the convergence of the iterative process are analyzed. It is proved that convergence can always be obtained. The added degree of freedom provides advantages over the classical algorithm in terms of convergence or BER.

Power extrinsic propagation for turbo-codes

Florence Alberge*, Ziad Naja and Pierre Duhamel,

Univ. Paris-Sud, UMR8506 Orsay, F-91405; CNRS, Gif-sur-Yvette, F-91192;

Supelec, Gif-sur-Yvette, F-91192, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette cedex, France

Tel: +33 1 69851757; fax: +33 1 69851769

e-mail:{alberge, naja, pierre.duhamel}@lss.supelec.fr

EDICS: SPC-CODC

Abstract

A generalized information passing strategy is discussed for iterative decoding of turbo-like systems. Both the existence of fixed point and the convergence of the iterative process are analyzed. It is proved that convergence can always be obtained. The added degree of freedom provides advantages over the classical algorithm in terms of convergence or BER.

Index Terms

Iterative Decoding, Maximum Likelihood Decoding, BICM, Turbo-Codes.

I. Introduction

Combined with iterative decoding, turbo codes can achieve near-Shannon-limit performance over both AWGN and Rayleigh fading channels [1]. These turbo-decoders were not originally introduced as the solution to an optimization problem, thus rendering their precise structure somewhat ad hoc and the convergence and stability analysis very difficult. As an example, forwarding extrinsic information between

both constituents of the receiver rather than a posteriori probabilities was not initially well understood. Among the various attempts to provide an analysis of iterative decoding, early results have been obtained with large block-length [2],[3] or when the corresponding graph is a tree [4]. In [5],[6] the turbo-decoding is interpreted as a dynamical system, leading to new but incomplete results. The failure to obtain complete results is mainly due to the inability to efficiently describe extrinsic information passing. A link between iterative decoding and optimization algorithms has been made recently in [7] where the turbo decoding is interpreted as the solution to a constrained optimization problem. Inspired by [7], we proposed in [8], [9] an alternative interpretation of iterative decoding. Similarities can been pointed out between these contributions. In both papers, iterative decoding is obtained as a relaxation of sequence-wise maximum likelihood. The dynamics are viewed as Gauss-Seidel iterations allowing a convergence analysis based on the results of Moré in [10]. Among the differences, the relaxation in [8], [9] provides a full understanding for extrinsic propagation and introduces naturally a new degree of freedom (β to be defined in section III) in information passing. Standard iterative decoding can be recovered as a special case ($\beta = 1$). This letter focuses on the role of β on the convergence. In [7], [5], conditions for convergence are provided assuming standard message passing ($\beta = 1$). In this letter, we prove that, for any given system and channel conditions, convergence can be obtained by choosing a modified message-passing algorithm, which can offer improved performance (faster convergence, smaller BER, ...). The existence of fixed-point is addressed by extending the proofs in [5] to encompass the general case. The results in this paper focus on iterative decoding for Bit Interleaved Coded Modulation (BICM) and for Serially Concatenated Turbo-Codes (SCTC). They can however be applied to a large class of problems.

II. Maximum likelihood decoding

Consider a typical communication scheme where a binary message $\mathbf{b}$ is encoded and sent over a memoryless channel to create the received data $\mathbf{y}$. Without any prior information on the input, the

maximum likelihood sequence detection (MLSD) reads

$$\widehat{\mathbf{b}}_{MLD} = \arg \max_{\mathbf{b}\in\{0,1\}^{n_b}} p(\mathbf{y} \mid \mathbf{b}) \tag{1}$$

where $p(\mathbf{y} \mid \mathbf{b})$ is the likelihood function depending on the transmission scheme and n_b is the length of $\mathbf{b}$. This is applied below to a BICM system depicted on Fig. 1. The sequence $\mathbf{b}$ is first encoded by a convolutional encoder to produce the sequence $\mathbf{c}$ of length n. Let $\mathbf{d} = \pi(\mathbf{c})$ denote the interleaved sequence. The complex transmitted signal s_k, $1 \leq k \leq n/m$, is chosen from an M-ary constellation ψ where ψ denotes the mapping scheme ($M = 2^m$) and sent over a noisy memoryless channel which provides the channel outputs y_k. The SCTC encoder [11] is also depicted in Fig. 1 where the mapping sub-block is replaced by a second encoder. For both systems, there is a one to one correspondence between $\mathbf{b}$ and $\mathbf{d}$. The MLSD problem in (1) is equivalent to searching $\widehat{\mathbf{d}}_{MLD}$ as:

$$\widehat{\mathbf{d}}_{MLD} = \arg \max_{\mathbf{d}\in\{0,1\}^{n}} \widetilde{p}_1(\mathbf{d})\widetilde{p}_2(\mathbf{d}) \tag{2}$$

where $\widetilde{p}_1(\mathbf{d})$ is the probability of receiving $\mathbf{y}$ when the sequence transmitted through the channel is either the mapping of $\mathbf{d}$ (for BICM) or the output of the inner code (for SCTC) and where $\widetilde{p}_2(\mathbf{d})$ is the indicator function of the outer code. A practical implementation of this optimization problem is infeasible. In practice, a turbo-like process is applied at the receiver side. The relationship between the MLSD in (2) and extrinsic propagation in turbo-codes was not originally well understood. Decisive steps have been made in [7] and later on in [8] in order to provide a complete interpretation of the turbo-like process as a distributed optimization strategy derived from the optimal MLSD. It is proved in [8] that turbo-codes aim at finding a solution to the n optimization problems below:

$$\left(\hat{\ell}_k, \hat{q}_k\right) = \arg \max_{\ell_k q_k \in \mathcal{F}} \tilde{\mathcal{C}}_k(\ell, q) \quad 1 \leq k \leq n \tag{3}$$

$$\tilde{\mathcal{C}}_k(\ell, q) \stackrel{def}{=} \sum_{d_k} \left(\sum_{\mathbf{d}:d_k} \widetilde{p}_1(\mathbf{d}) \prod_i q_i(d_i) \right) \left(\sum_{\mathbf{d}':d_k} \widetilde{p}_2(\mathbf{d}') \prod_i \ell_i(d'_i) \right) \tag{4}$$

where $\mathcal{F}$ is the set of all possible PMFs on d_k. There are two major differences between the sequence-wise maximum likelihood decoding in (2) and the suboptimal decoding in (3-4). First, the discrete variable

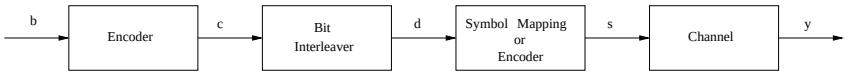


Figure 1. BICM / SCTC - Transmission schemes

$\mathbf{d}$ in (2) has been replaced by the continuous variable $p(\mathbf{d}) = \prod_i \ell_i(d_i) q_i(d_i)$. This does not correspond to an approximation [7],[8], since the optimal solution is not changed. Then the full concordance is relaxed to a single-bit concordance in (3-4). Indeed, in products $\widetilde{p}_1(\mathbf{d})\widetilde{p}_2(\mathbf{d}')$ involved in (4), $\mathbf{d}$ and $\mathbf{d}'$ are such that $d_k = d'_k$ whereas $\mathbf{d} = \mathbf{d}'$ in (2). Based on this framework, a game-theoretic interpretation of iterative decoding has been given in [9]. Iterative decoding can be seen as a game involving n players (the individual bits d_k), each player attempting to maximize its own utility function $\widetilde{\mathcal{C}}_k$. The efficiency of the joint optimization process can be evaluated through the social welfare of the game which is defined as the sum of the utilities of all players [12] $\widetilde{\mathcal{C}} = \sum_{k=1}^{n} \widetilde{\mathcal{C}}_k$. The connection between $\widetilde{\mathcal{C}}$ and the MLSD is detailed in [8]. The next section provides an iterative scheme for the maximization of $\widetilde{\mathcal{C}}$.

III. Iterative Decoding

Let $f_{d_k}(u_{-k}, \widetilde{p}_\alpha)$ be defined as $f_{d_k}(u_{-k}, \widetilde{p}_\alpha) = \sum_{\mathbf{d}:d_k} \widetilde{p}_\alpha(\mathbf{d}) \prod_{j \neq k} u_j(d_j)$. Then, (3-4) reads

$$\left(\hat{\ell}_k, \hat{q}_k\right) = \arg \max_{\substack{\ell_k q_k \\ \ell_k, q_k \in \mathcal{F}}} \sum_{d_k} \ell_k(d_k) q_k(d_k) f_{d_k}(q_{-k}, \widetilde{p}_1) f_{d_k}(\ell_{-k}, \widetilde{p}_2) \tag{5}$$

with solution $\widehat{\ell}_k(d_k)\widehat{q}_k(d_k) = 1$ if $f_{d_k}(q_{-k}, \widetilde{p}_1) f_{d_k}(\ell_{-k}, \widetilde{p}_2) > f_{\overline{d}_k}(q_{-k}, \widetilde{p}_1) f_{\overline{d}_k}(\ell_{-k}, \widetilde{p}_2)$ and 0 otherwise, where $\overline{d}_k = 1 - d_k$. An iterative process propagating hard estimates is likely to get stuck in a local minimum. Therefore, the optimization problem in (5) is solved by propagating soft estimates [8], [9]:

$$\widehat{\ell}_k(d_k)\widehat{q}_k(d_k) \propto \left(f_{d_k}(q_{-k}, \widetilde{p}_1) f_{d_k}(\ell_{-k}, \widetilde{p}_2)\right)^{\beta} \tag{6}$$

where β is a positive constant. We can observe that $\beta \to \infty$ (combined with a normalization step) yields the hard estimates. The soft estimates in (6) can alternately be understood as the solution to a regularized

optimization problem:

$$\left(\hat{\ell}_k, \hat{q}_k\right) = \arg \max_{\substack{\ell_k q_k \\ \ell_k, q_k \in \mathcal{F}}} \beta \sum_{d_k} \ell_k(d_k) q_k(d_k) \log\left(f_{d_k}(q_{-k}, \widetilde{p}_1) f_{d_k}(\ell_{-k}, \widetilde{p}_2)\right) + H\left(\ell_k(d_k) q_k(d_k)\right) \quad (7)$$

The maximization of the first term of (7) gives the same hard solution than (5). The second term, $H\left(\ell_k(d_k) q_k(d_k)\right)$, is the entropy of $\ell_k(d_k) q_k(d_k)$. It is interesting to note that the metric in (7) takes the form of a free energy, which parallels another alternative interpretation for iterative decoding that observes a connection with Bethe/Kikuchi approximations in statistical physics [13], [14], [15] though we will not consider this connection in this correspondence. Equation (6) gives a characterization of the product $\widehat{\ell}_k \widehat{q}_k$. The individual values of $\widehat{\ell}_k$ and $\widehat{q}_k$ depend on the scheduling of the successive updates applied to the rule (6). A Jacobi/Gauss-Seidel implementation gives the following updates [8]:

$$q_k^{(it)}(d_k) \propto \left(\sum_{\mathbf{d}: d_k} \widetilde{p}_2(\mathbf{d}) \prod_{j \neq k} \ell_j^{(it-1)}(d_j)\right)^{\beta} \quad (8)$$

$$\ell_k^{(it)}(d_k) \propto \left(\sum_{\mathbf{d}: d_k} \widetilde{p}_1(\mathbf{d}) \prod_{j \neq k} q_j^{(it)}(d_j)\right)^{\beta} \quad (9)$$

In the coding community, $\ell_k(d_k)$, $q_k(d_k)$ are usually called *extrinsics* and $\ell_k(d_k) q_k(d_k)$ (after normalization) is the *APP* (*A Posteriori Probability*). As a result, extrinsic propagation proceeds from a distributed maximization of $\widetilde{C}$ with respect to the APP ($\ell_k(d_k) q_k(d_k)$). The separation into extrinsics (ℓ_k, q_k) occurs exclusively in the dynamics and proceeds from a numerical solution method. In the literature, iterative decoding is (8-9) with $\beta = 1$. This is inherited from the seminal paper of Berrou [1] and may be sub-optimal. This letter analyses the role of β in the convergence of the iterative process and the existence of solutions to the fixed point problem in (8-9). Classical turbo-decoding ($\beta = 1$) algorithms always possess (at least) one fixed point [5]. We prove here that this property can be extended to a wider subset. The fixed-point equations in (8-9) read:

$$\lambda_{\ell,k} + \lambda_{q,k} = \beta[\Pi_{\widetilde{p}_1}(\lambda_q)]_k - (\beta - 1)\lambda_{q,k} \quad 1 \leq k \leq n \quad (10)$$

$$\lambda_{\ell,k} + \lambda_{q,k} = \beta[\Pi_{\widetilde{p}_2}(\lambda_\ell)]_k - (\beta - 1)\lambda_{\ell,k} \quad 1 \leq k \leq n \quad (11)$$

where $\lambda_{\ell,k} = \log\left(\frac{\ell_k(d_k)}{\ell_k(\overline{d}_k)}\right)$ and $\lambda_{q,k} = \log\left(\frac{q_k(d_k)}{q_k(\overline{d}_k)}\right)$ are log-likelihood ratios of the extrinsics and where $\Pi_{\widetilde{p}_1}(\lambda_q)$ and $\Pi_{\widetilde{p}_2}(\lambda_\ell)$ denote two LLR vectors with taps $[\Pi_{\widetilde{p}_1}(\lambda_q)]_k = \log\left(\frac{\sum_{\mathbf{d}:d_k}\widetilde{p}_1(\mathbf{d})\prod_j q_j(d_j)}{\sum_{\mathbf{d}:\overline{d}_k}\widetilde{p}_1(\mathbf{d})\prod_j q_j(d_j)}\right) = \log\left(\frac{f_{d_k}(q_{-k},\widetilde{p}_1)}{f_{\overline{d}_k}(q_{-k},\widetilde{p}_1)}\right) + \lambda_{q,k}$ and $[\Pi_{\widetilde{p}_2}(\lambda_\ell)]_k = \log\left(\frac{\sum_{\mathbf{d}:d_k}\widetilde{p}_2(\mathbf{d})\prod_j \ell_j(d_j)}{\sum_{\mathbf{d}:\overline{d}_k}\widetilde{p}_2(\mathbf{d})\prod_j \ell_j(d_j)}\right) = \log\left(\frac{f_{d_k}(\ell_{-k},\widetilde{p}_2)}{f_{\overline{d}_k}(\ell_{-k},\widetilde{p}_2)}\right) + \lambda_{\ell,k}$. In the sequel, we also use the shorthand notation $F_{\widetilde{p},\beta}(\lambda_u) = \beta\Pi_{\widetilde{p}}(\lambda_u) - (\beta-1)\lambda_u$.

Proposition 1: For any PMF $\widetilde{p}$, there always exists $\beta_{NE} \geq 1$ such that $\forall\beta \leq \beta_{NE}$ the map $F_{\widetilde{p},\beta}$ is a homeomorphism.

Proof: It is proved in [5] that there exists a constant c such that $\parallel \Pi_{\widetilde{p}}(x) - x \parallel \leq c$ for all x in $\mathbb{R}^n$. This is equivalent to $\parallel \beta\Pi_{\widetilde{p}}(x) - \beta x \parallel \leq \beta c$ with $\beta \in]0,+\infty[$ leading to

$$\parallel F_{\widetilde{p},\beta}(x) - x \parallel \leq c' \quad \forall x \in \mathbb{R}^n \tag{12}$$

Let J_F denote the Jacobian matrix of $F_{\widetilde{p},\beta}(x)$ and let J_Π denote the Jacobian matrix of $\Pi_{\widetilde{p}}(x)$. For any PMF $\widetilde{p}$ and for any $x \in \mathbb{R}^n$, J_Π is similar to a symmetric positive definite matrix S *ie* $D_\Pi^{-1/2} J_\Pi D_\Pi^{1/2} = D_\Pi^{-1/2} S D_\Pi^{1/2}$ [5]. From the definition of $F_{\widetilde{p},\beta}$, we have $J_F = \beta J_\Pi - (\beta-1)I$ where I is the $n \times n$ identity matrix. Thus J_F is similar to $\beta S - (\beta-1)I$. If $\beta \geq 1$, $\beta S - (\beta-1)I$ is a positive definite matrix (this is true for any PMF $\widetilde{p}$ and for any $x \in \mathbb{R}^n$). For $\beta > 1$ the positive definiteness of J_F depends of the eigenvalues of $\Pi_{\widetilde{p}}(x)$. Thus, there always exists $\beta_{NE} \geq 1$ such that $\forall\beta \leq \beta_{NE}$, J_F is positive definite which leads to $det(J_F) > 0$ for all $x \in \mathbb{R}^n$, it follows that $F_{\widetilde{p},\beta}$ is locally invertible and that it has a continuous inverse. Since $F_{\widetilde{p},\beta}$ is onto and one-to-one (12), $F_{\widetilde{p},\beta}$ is a homeomorphism. ■

We are now ready to prove the existence of fixed point in the general case.

Theorem 1: There always exists $\beta_{NE} \geq 1$ such that $\forall\beta \leq \beta_{NE}$, at least one solution exists to (10-11).

Proof: Based on the proofs in [5], we define the map $\theta : \mathbb{R}^n \mapsto \mathbb{R}^n$ by $\theta(y) = (y - F^{-1}_{\widetilde{p}_1,\beta}(y)) + (y - F^{-1}_{\widetilde{p}_2,\beta}(y))$. If (λ_l, λ_q) solves (10) and (11) then $y = \lambda_l + \lambda_q$ is a fixed-point of θ. Conversely, if y is a fixed point of θ then $\lambda_l = y - F^{-1}_{\widetilde{p}_2,\beta}(y)$ and $\lambda_q = y - F^{-1}_{\widetilde{p}_1,\beta}(y)$ solves (10) and (11). From (12), we obtain that there exists a constant c such that $\parallel \Theta(y) \parallel \leq c$. Thus, from the Brouwer fixed point theorem

we obtain that the map $y \mapsto \Theta(y)$ possesses a fixed point. ■

The existence of (at least) one solution of (10-11) has been proved when $\beta \in]0, \beta_{NE}]$ with $\beta_{NE} \geq 1$ and depends of the two PMFs $\widetilde{p}_1$, $\widetilde{p}_2$. Even if the existence of a fixed-point is proved, there is no guarantee that the iterative procedure will converge. This point is addressed below. We use the compact notation $S_\beta \lambda = 0$ to denote the system of equations (10-11) with $\lambda = [\lambda_\ell \ \ \lambda_{\mathbf{q}}]^T$. Let $\mathbf{C}_\alpha$, $\alpha \in \{1, 2\}$, be matrices with elements $[\mathbf{C}_\alpha]_{i,j} = p_\alpha[d_j = 1 | d_i = 1] - p_\alpha[d_j = 1 | d_i = 0]$ where $p_1(\mathbf{d}) = K_1 \widetilde{p}_1(\mathbf{d}) \mathbf{q}(\mathbf{d})$ and $p_2(\mathbf{d}) = K_2 \widetilde{p}_2(\mathbf{d}) \ell(\mathbf{d})$ and K_1, K_2 are normalization factors. The Jacobian ∇S_β reads [7]:

$$\nabla S_\beta = \begin{pmatrix} \mathbf{I} & \beta(\mathbf{I} - \mathbf{C}_1) \\ \beta(\mathbf{I} - \mathbf{C}_2) & \mathbf{I} \end{pmatrix} \tag{13}$$

where $\mathbf{I}$ is the identity matrix of size $n \times n$. Global convergence has been studied in [7]. In particular conditions are obtained for ∇S_β to be an M-matrix [16]. These conditions are not easily checkable in a practical setting. Here, the question of the convergence is addressed in a different way. We demonstrate below that, for any PMF $\widetilde{p}_1$ and $\widetilde{p}_2$, there are suitable values of β for which ∇S_β is a strictly diagonally dominant matrix. The proof is based on the work of Moré [10] on nonlinear Jacobi/Gauss-Seidel methods.

Theorem 2 (Threshold for convergence): Let $S_\beta : D \subset \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be continuous on D, and assume that $S_\beta \lambda = 0$ has a solution $\lambda^* \in D$, and that for some $r \geq 0$, $Q = \{\lambda \in \mathbb{R}^{2n} : \| \lambda - \lambda^* \|_\infty \geq r\} \subset D$. It always exists β_0 such that, $\forall \beta \leq \beta_0$, ∇S_β is a strictly diagonal dominant matrix, λ^* is unique in Q, and for any $\lambda^{(0)}$ in Q the Jacobi and Gauss-Seidel sequences are well-defined and converge to λ^*.

Proof: The continuity of S_β has been proven in proposition 1. Existence of solutions for $S_\beta \lambda = 0$ in $\mathbb{R}^{2n}$ has been proven when $\beta \leq \beta_{NE}$. Let $\lambda_i^{(k)}$ denote the value of λ_i at iteration k. If $|\lambda_i^{(k)}| < +\infty$ then Q does exist. This is likely to be so in a noisy environment. It is proved in proposition 3 of [9] that it always exists some $\beta_C > 0$ such that, $\forall \beta \leq \beta_C$, ∇S_β is a strictly diagonal dominant matrix for any $\lambda \in \mathbb{R}^{2n}$. Choosing $\beta_0 = min(\beta_{NE}, \beta_C)$ and applying theorem 5.7 in [10] concludes the proof. ■

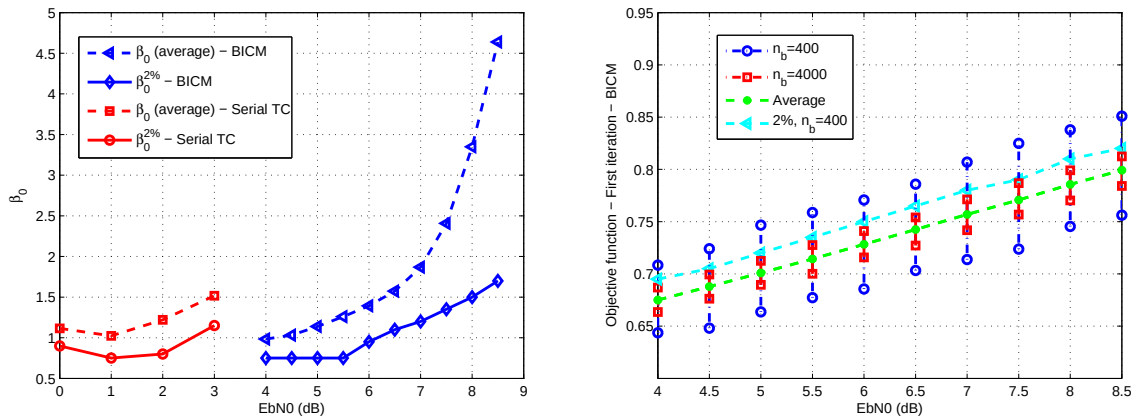


Figure 2. Left: $\overline{\beta_0}$ and $\beta_0^{2\%}$ vs $\frac{E_b}{N_0}$ for BICM ($n_b = 400$) and SCTC. Right: $K_m \tilde{\mathcal{C}}_{map}^{(1)}$ vs $\frac{E_b}{N_0}$ for BICM

The choice of β is a trade-off between convergence and optimality of the joint-detection process. Small values of β means convergence of the iterative process (maybe towards a local minima). High values of β means high values of the social welfare $\tilde{\mathcal{C}}$ [8] but the iterative process may not converge.

IV. Simulation

The proposed algorithm is tested in both BICM and SCTC cases. The BICM scheme involves a $(5,7)_8$ convolutional code of rate $1/2$. Bits are mapped to 16-QAM using Set Partitioning. The number of information bits is either $n_b = 400$ or $n_b = 4000$. The signal to noise ratio is defined as $\frac{E_b}{N_0}$, where E_b denotes the energy per information bit and N_0 the noise variance. We also consider a serially-concatenated turbo-code (SCTC) with a $(5,7)_8$ outer code. The inner code is a convolutional code with generator $\frac{1}{1+D}$, $n_b = 250$. First check the new degrees of freedom. The threshold between convergence/non-convergence of the iterative process (i.e. β_0 in theorem 2) is a variable of interest. β_0, depending on the received sequence is a random variable. Its average value, $\overline{\beta_0}$, is plotted as a function of $\frac{E_b}{N_0}$ in Fig 2, as well as $\beta_0^{2\%}$ defined as follows: $Pr(\beta_0 \leq \beta_0^{2\%}) = 2\%$. Both $\beta_0^{2\%}$ and $\overline{\beta_0}$ are estimated over 4000 runs for each $\frac{E_b}{N_0}$. We observe the following. Both $\overline{\beta_0}$ and $\beta_0^{2\%}$ are increasing functions of $\frac{E_b}{N_0}$ at least in waterfall and

error-floor regions. From Fig. 3 and 4, the waterfall region is $[1.5dB; 3dB]$ for SCTC and $[5dB; 6.5dB]$ for BICM. In the waterfall region and with standard iterative decoding, misconvergence behaviors are likely to be observed ($1 \leq \overline{\beta_0} \leq 1.5$). Thus, β should be decreased to allow convergence. In the error-floor region ($\frac{E_b}{N_0} \geq 7dB$ for BICM), both $\overline{\beta_0}$ and $\beta_0^{2\%}$ are significantly above 1. Misconvergence is expected to be a rare phenomenon and β can be increased to improve the convergence rate. Based on these observations, the following strategies are proposed. First consider BICM. Extra simulations (not reported) show that decreasing β in the waterfall region induces convergence but does not improve the BER. Hence, the standard value $\beta = 1$ is maintained at low $\frac{E_b}{N_0}$. For $\frac{E_b}{N_0} \geq 6.5dB$, $\beta_0^{2\%}$ is a linearly increasing function of $\frac{E_b}{N_0}$. The range of variations of $K_m\tilde{\mathcal{C}}_{map}^{(1)}$ is plotted in Fig 2 where $\tilde{\mathcal{C}}_{map}^{(1)}$ is $\tilde{\mathcal{C}}$ evaluated at the output of the demapping sub-block at the first iteration and K_m is a normalization factor such that $K_m\tilde{\mathcal{C}}_{map}^{(1)}$ span $[0\ ;\ 1]$. The same definition holds for $K_c\tilde{\mathcal{C}}_{cod}^{(k)}$ at the output of the decoder and at iteration k. Let $\overline{\tilde{\mathcal{C}}_{map}^{(1)}}$ denote the average value of $\tilde{\mathcal{C}}_{map}^{(1)}$ and let $\tilde{\mathcal{C}}^{2\%}$ be defined such that $Pr(K_m\tilde{\mathcal{C}}_{map}^{(1)} \geq \tilde{\mathcal{C}}^{2\%}) = 2\%$. Both $\overline{\tilde{\mathcal{C}}_{map}^{(1)}}$ and $\tilde{\mathcal{C}}^{2\%}$ are linearly increasing function of $\frac{E_b}{N_0}$. Thus the instantaneous value of $K_m\tilde{\mathcal{C}}_{map}^{(1)}$ provides a (rough) estimation of $\frac{E_b}{N_0}$ and an accurate value for β can be associated. Based on the discussion above, the value of β is adjusted in the first iteration with the rule $\beta = max(1; a_1 K_m\tilde{\mathcal{C}}_{map}^{(1)} + a_0)$ where a_0 and a_1 are chosen such that $a_1\tilde{\mathcal{C}}^{2\%} + a_0$ provides the better fit, in the least square sense, to $\beta_0^{2\%}$ for $\frac{E_b}{N_0} \in [6dB\ ;\ 8.5dB]$. With BICM and $n_b = 400$, $a_0 = -5.8$, $a_1 = 8.8$. A classical BICM system makes use of the iterative process (8-9) with $\beta = 1$. In general, the iterative process is stopped either when the maximum number of iterations is reached or when a convergence rule is met (stopping criteria A in table I). We define another stopping criterion (denoted as B in table I) which stops also when an oscillating behaviour [17] is detected (useful at low $\frac{Eb}{N_0}$) or when a target value of $\tilde{\mathcal{C}}$ is reached (useful at high $\frac{Eb}{N_0}$). The rule used to compute the target value is based on the joint observation that both the target value and $K_m\tilde{\mathcal{C}}_{map}^{(1)}$ are linear functions of $\frac{Eb}{N_0}$. The classical BICM ($\beta = 1$, stopping criterion A) is compared to a BICM with optimized parameters ($\beta = max(1, a_1\tilde{\mathcal{C}}_{map}^{(1)} + a_0)$, stopping criterion B).

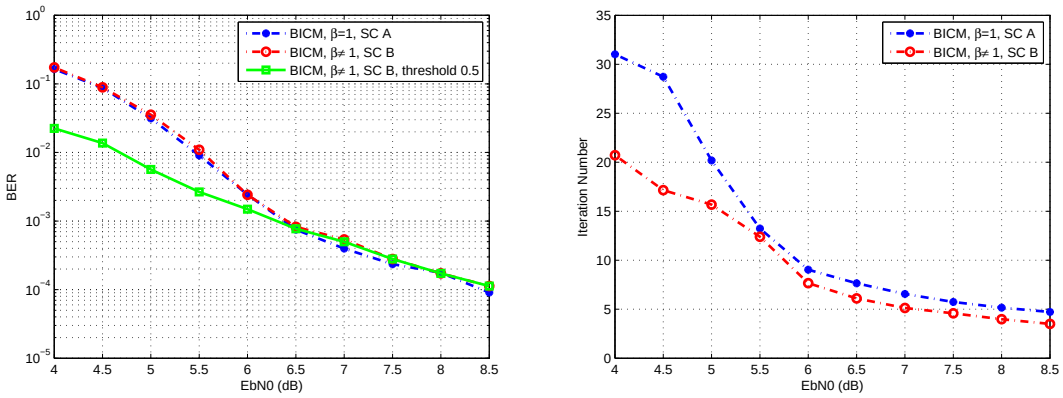


Figure 3. BICM, $n_b = 400$: BER vs $\frac{Eb}{N_0}$ (left), Average number of iterations vs $\frac{Eb}{N_0}$ (right)

Stopping criterion A	$\lvert K_c\tilde{\mathcal{C}}_{cod}^{(k)} - K_m\tilde{\mathcal{C}}_{map}^{(k)}\rvert \leq 10^{(-5)} k \geq it_{max}$ **or** $it_{max} = \begin{cases} 50 \text{ for BICM} \\ 10 \text{ for SCTC} \end{cases}$
Stopping criterion B (BICM)	Stopping criterion A **or** 3 oscillations detected (slope of the objective function: sign detection) **or** $K_c\tilde{\mathcal{C}}_{cod}^{(k)} > K_m\tilde{\mathcal{C}}_{map}^{(1)} + 0.1$ (Target value)

Table I

DESCRIPTION OF STOPPING CRITERION A (CLASSICAL) AND B

The BER and the number of iterations are plotted in Fig. 3. The stopping criterion B combined with an adaptive value of β always gives a smaller number of iterations. We can see that, for an $\frac{E_b}{N_0}$ above 7dB, the improvement is mainly due to the adaptive value of β (the oscillation detector is almost always inactive). At the opposite, β is equal to 1 at low $\frac{E_b}{N_0}$ and the improvement is mainly due to the oscillation detector. Another opportunity is provided by this new setting: since the approximation between MLD and the optimisation criterion (social welfare) is now clear, the value obtained by the social welfare at convergence gives an indication on how close we are from MLD solutions. For that purpose, we plot in Fig. 3 the BER corresponding to frames with $K_c\tilde{\mathcal{C}}_{cod}^{(\infty)}$ above 0.5 suggesting that above a certain $\frac{E_b}{N_0}$, the solution is always in the vicinity of the MLD. Numerical results showing the behavior of $\tilde{\mathcal{C}}$ are also

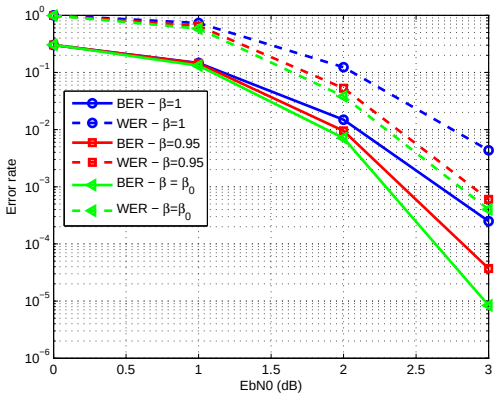


Figure 4. SCTC: BER and WER vs EbN0 for various values of the additional parameter β

reported in [7],[13]. For SCTC, the BER and word-error rate (WER) are compared when $\beta = 1$ and $\beta = 0.95$. The BER and WER are also plotted for the optimal situation $\beta = \beta_0$. This is not a practical situation since β_0 is estimated by running several times the iterative process, but this curve is a lower bound. We observe that choosing $\beta = 0.95$ in the iterative process instead of $\beta = 1$ improves both the BER and WER and is close to the lower bound for the WER.

V. Conclusion

A generalized standard message-passing is provided for turbo-codes. It is proved that convergence is attainable by propagating extrinsics raised to some power, which constitutes an additional degree of freedom compared to the classical situation. Performance can therefore be improved, as illustrated for SCTC and BICM. The evaluation of the objective function $\tilde{\mathcal{C}}$ allows well-informed decisions at the receiver's side: *(i)* at first iteration by guiding the choice of the power β, *(ii)* during the iterations by deciding of early stop, and *(iii)* at convergence by providing indications on the reliability of the solution.

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error-correcting coding and decoding: Turbo codes,” in *Proc. IEEE Int. Conf. Commun*, 1993, pp. 1064–1070.

[2] H. E. Gamal and A. Hammons, “Analysing the turbo decoder using the Gaussian approximation,” *IEEE Trans. on Inform. Theory*, vol. 47, pp. 671–686, Feb. 2001.

[3] S. ten Brink, “Convergence behavior of iteratively decoded parallel concatenated codes,” *IEEE Trans. on Commun.*, vol. 49, pp. 1727–1737, Oct 2001.

[4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.

[5] T. Richardson, “The geometry of turbo-decoding dynamics,” *IEEE Trans. on Inform. Theory*, vol. 46, no. 1, pp. 9–23, 2000.

[6] S. Ikeda, T. Tanaka, and S. Amari, “Information geometry of turbo and low-density parity-check codes,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1097–1114, Jun. 2004.

[7] J. M. Walsh, P. Regalia, and C. R. Johnson, “Turbo decoding as Iterative Constrained Maximum-Likelihood Sequence Detection,” *IEEE Trans. on Inform. Theory*, vol. 52, pp. 5426–5437, Dec. 2006.

[8] F. Alberge, Z. Naja, and P. Duhamel, “From Maximum Likelihood to Iterative Decoding,” in *ICASSP Proc.*, Prague, Czech Republic, 22-27 May 2011.

[9] F. Alberge, “A game-theoretic interpretation of iterative decoding,” in *EUSIPCO*, vol. 1, Barcelona, Spain, 29 Aug. - 2 Sept. 2011, pp. 76–80.

[10] J. Moré, “Nonlinear generalizations of matrix diagonal dominance with application to Gauss-Seidel iterations,” *SIAM J. Numer. Anal.*, vol. 9, no. 2, pp. 357–378, 1972.

[11] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, “Serial concatenation of interleaved codes: performance analysis, design and iterative decoding,” in *IEEE International Symposium on Information Theory*, 1997.

[12] S. Lasaulce, M. Debbah, and E. Altman, “Methodologies for Analysing Equilibria in Wireless Games,” *IEEE Signal Proc. Magazine*, vol. 26, no. 5, pp. 41–52, 2009.

[13] J. M. Walsh and P. A. Regalia, “On the relationship between belief propagation decoding and joint maximum likelihood detection,” *IEEE Trans on Commun.*, vol. 58, no. 10, pp. 2753–2758, 2010.

[14] A. Montanari and N. Sourlas, “The statistical mechanics of turbo-codes,” *Eur. Phys. J. B.*, no. 18, p. 107, 2000.

[15] J. Yedidia, W. Freeman, and Y. Weiss, “Constructing free-energy approximations and generalized belief propagation algorithms,” *IEEE Transactions on Information Theory*, no. 7, pp. 2282–2312, Jul. 2005.

[16] R. Horn and C. Jonhson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[17] L. Kocarev, F. Lehmann, G. Maggio, B. Scanavino, Z. Tasev, and A. Vardy, "Nonlinear dynamics of iterative decoding systems: analysis and applications," *IEEE Trans. on Infor. Theory*, vol. 52, no. 4, pp. 1366–1384, 2006.

Annexe 5

F. Alberge. On some Properties of the Mutual Information between Extrinsics with Application to Iterative Decoding. IEEE Trans. on Communications, vol 63, n° 5, 2015.

Abstract

Iterative decoding is an efficient error-correction tool based on the exchange of extrinsic probabilities between the constituent decoders. In this paper, the properties of the mutual information between the extrinsic LLR at the output of two constituent decoders are analyzed with application to turbo and LDPC codes. This is a bridge between information-theoretic analysis and practical implementations. It is proved here that the mutual information between extrinsics is a lower bound of the mutual information between each extrinsic and the transmitted message. In addition, an efficient online evaluation is provided in the paper with accuracy validated through numerical experiments. As an application, the mutual information between extrinsics is used for designing efficient stopping criterion and error detection rules at the decoder side. Two online methods for the estimation of optimal scaling factor to be applied to the extrinsic LLR are also derived. In contrast with most references, an analytical expression is obtained that does not require estimation of the actual transmitted bits. All results in the paper are derived for Gaussian distributed LLR with independent mean and variance.

On some Properties of the Mutual Information between Extrinsics with Application to Iterative Decoding

Florence Alberge
LSS (UMR8506), Univ. Paris-Sud, CNRS, CentraleSupelec,
3 rue Joliot-Curie, 91192 Gif-sur-Yvette cedex, France
Tel: +33 1 69851757; fax: +33 1 69851765 ; e-mail:alberge@lss.supelec.fr

***Abstract*—Iterative decoding is an efficient error-correction tool based on the exchange of extrinsic probabilities between the constituent decoders. In this paper, the properties of the mutual information between the extrinsic LLR at the output of two constituent decoders are analyzed with application to turbo and LDPC codes. This is a bridge between information-theoretic analysis and practical implementations. It is proved here that the mutual information between extrinsics is a lower bound of the mutual information between each extrinsic and the transmitted message. In addition, an efficient online evaluation is provided in the paper with accuracy validated through numerical experiments. As an application, the mutual information between extrinsics is used for designing efficient stopping criterion and error detection rules at the decoder side. Two online methods for the estimation of optimal scaling factor to be applied to the extrinsic LLR are also derived. In contrast with most references, an analytical expression is obtained that does not require estimation of the actual transmitted bits. All results in the paper are derived for Gaussian distributed LLR with independent mean and variance.**

***Index Terms*—Mutual Information, Iterative Decoding, Stopping Rule, EXIT Charts.**

I. Introduction

Extrinsic information transfer (EXIT) charts were first introduced in [1] for the analysis of iterative decoding. The principle is quite general and has been applied to turbo-codes [2], LDPC codes [3] or repeat-accumulate codes [4] and extended to non-binary iterative decoding in [5]. In EXIT analysis, the mutual information $I(L, X)$ between the extrinsic log-likelihood ratios (LLR) L and the binary message X is computed for performance evaluation. This is an offline tool which allows to compute a performance rating on the reliability of a solution and is usually intended to design capacity-approaching codes before implementing them. In an actual transmission, the receiver decodes the transmitted message via an iterative turbo decoding process based on extrinsic propagation [6]. In this iterative receiver, the extrinsic information is updated within each individual decoding block and passed to the other decoder over many iterations. This is a pragmatic approach since iterative decoding was not originally introduced as the solution to an optimization problem rendering its analysis difficult. As a consequence, the stopping criterion of the iterative process can not easily be derived from the evaluation of an objective function which would converge to some limit value. In addition, the mutual information $I(L, X)$ which is the performance indicator in the Exit Charts is not easily computable at the receiver. Thus, different classes of approximations and rules have been developed for error detection and stopping criteria. One important class of stopping criteria is concerned with the detection of a stationary point. This class encompasses cross entropy [7] stopping rule, sign change rule [8], [9] or hard-decision-aided [10] rule as special cases. These rules have difficulties stopping the decoding if the iterative process enters an oscillatory behavior [11] and do not have any error detection capability. Another class is based on a performance evaluation like in [12] where the stopping rule is based on the mean of the absolute values of the log-likelihood ratios at the output of the component decoders or on the instantaneous values as in [13]. It is well-known that the reliability of a solution is connected with the magnitude of the LLR however the choice of the threshold is not trivial.

Independently, several publications such as [14], [15], [16] considered the issue of optimal LLR scaling. Indeed, it was frequently observed in turbo-codes and LDPC codes that the imperfections in the receiver can substantially lower the performance of the decoder. Linear LLR correction can potentially compensate the degradation. The LLR-distribution is analyzed for both LDPC and turbo-decoder in [17], [18]. It is pointed out that the LLR are well approximated by Gaussian distributions but it is inappropriate to use the mean of the density only to model the iterative decoding process. Using this result, the LLR are modelled in this paper using Gaussian distribution including a scaling factor which is equivalent to considering independent mean and variance.

This paper shows that the two issues (stopping rule and optimal scaling) can be addressed jointly by introducing a new performance indicator. This indicator is the mutual information between extrinsics $I(L_y, L_z)$ where L_y and L_z denote the extrinsic LLRs at the output of two individual component decoders. The contributions of the paper are listed below. The mutual information between extrinsics is proved to be a lower bound of the mutual information between the extrinsics LLR and the message X to be retrieved. As a consequence, $I(L_y, L_z)$ is an indicator for a successful/unsuccessful decoding and a good metric for stopping criterion. The mutual information between extrinsics can be measured offline via an histogram method as in Exit Charts. A single curve Exit

Chart [19] is used here since this configuration is suitable for tracking the evolution of a given system as a whole and enables an easy evaluation of $I(L_y, L_z)$. In parallel, an efficient online computation of $I(L_y, L_z)$ is derived. Comparison with the results of the Exit Charts shows a good accuracy of the online estimation. In addition to the previous results, two methods are proposed for the online estimation of the optimal scaling factor. These methods allow closed-form expression and do not require an estimation of the actual message bits as in [16], [20]. It should be mentioned however that the latter references do not require Gaussian assumption and are more general than the methods proposed here. It will be shown, through simulations that, in the context of Gaussian LLR, the methods proposed here outperform [16] in particular in early iterations and at low SNR. The connection between the optimal scaling factor and the mutual information between extrinsics is also emphasized. It is well understood that the optimal scaling factor depends on the SNR, the iteration number and the code structure. The evolution of the mutual information between extrinsics shares the same characteristics. We will see in this paper that the optimal scaling factor can be viewed as a function of a single parameter: the mutual information $I(L_y, L_z)$ between extrinsics. The main contribution of the paper is thus to provide a theoretical analysis and a practical method for the full exploitation of the information contained in the extrinsics for a proper design of the receiver enabling self-diagnosis (successful decoding, early stop) and well-informed choices (scaling factor). The proposed methodology is quite general and rely partially on the assumption that the LLR are Gaussian distributed and that the symmetry property holds.

II. System model and notation

In this paper, random variables are denoted with upper-case letters and their corresponding realizations with lower-case letters. Sequences of random variables or vectors are indicated by boldface letters. Fig. 1 depicts the decoding model used in this paper and previously in [21]. The a priori channel models the a priori information used at the constituent decoders and the communication channel models the transmission medium between the transmitter and the receiver. Both serial and parallel turbo coding schemes can be seen as specific instantiations of this generic scheme. For the iterative decoding of an outer decoder in a serial concatenation, the switch in Fig. 1 is in position 1 and the communication channel is inactive. When the switch is in position 2, the system can be used for modeling the iterative decoding of a constituent decoder in a parallel concatenation setting. This system can also be used for modeling the input/output relationship at the Variable Node or at the Check Node decoder in LDPC. In both systems, the iterative decoding aims at finding the binary sequence $\mathbf{X} = (X_1, X_2, ..., X_K)$ with length K. A constituent decoder takes as input an a priori information $\mathbf{L_y}$ (resp. $\mathbf{L_z}$) which denotes log-likelihood ratios (LLR) of an extrinsic probability. The length of $\mathbf{L_y}$ (resp. $\mathbf{L_z}$) is K and $L_{y,k}$ is the k^{th} element of $\mathbf{L_y}$. At iteration i, the first decoder receives the prior information $\ell_y^{(i)}$ and outputs the regenerated prior information $\ell_z^{(i)}$ which comes as an input to the second

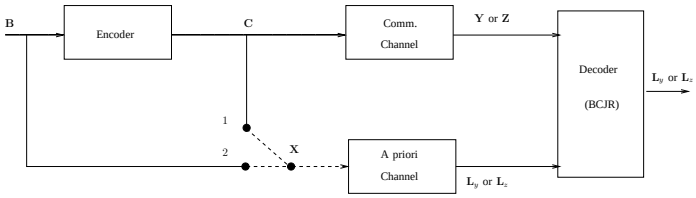


Figure 1. General Encoding/Decoding model

decoder and gives $\ell_y^{(i+1)}$ as an output. We then focus on the statistical properties of the extrinsic LLR. In other words, $L_{y,k}$ (or $L_{z,k}$) is considered as an outcome of random variable L_y (or L_z). In the following, index k will be omitted (except when needed for clarification) and we will use simply X, L_y, L_z or even L when an equation or a property hold for both L_y and L_z. Unless stated otherwise, the next properties hold [22]: **Symmetry**: $p_L(\ell|X) = p_L(-\ell| - X)$, **Generalized Consistency**: $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\alpha\ell}$ with $\alpha \in \mathbb{R}^+$, **Range**: $\ell \in]-\infty; +\infty[$. In the seminal works on EXIT Charts and density evolution [2], [19], [23], the assumptions considered are symmetry and consistency which corresponds to $\alpha = 1$ in the Generalized Consistency property above. This comes from the assumption that L is a noisy version of $\mathbf{X}$ with expression:

$$L = \frac{\sigma^2}{2}X + \sigma n \tag{1}$$

with $n \sim \mathcal{N}(0, 1)$. In that case, $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\ell}$. A stochastic analysis of iterative decoding is available in [17] where it is shown that the input-output signals in a turbo-decoder, when expressed using LLR, are indeed well approximated by a Gaussian distribution but with independent mean and variance. These results also hold for LDPC decoders [18]. In this paper, we will consider that

$$L = \alpha\frac{\sigma^2}{2}X + \sigma n \tag{2}$$

with $n \sim \mathcal{N}(0, 1)$. In that case, $\frac{p_L(\ell|X=1)}{p_L(-\ell|X=1)} = e^{\alpha\ell}$ and the Generalized consistency condition holds. This model is in accordance with the true LLR distribution that can be observed in iterative decoding of turbo or LDPC codes.

III. Mutual Information

The mutual information between extrinsics is analyzed first. The statistical model in (2) is considered and an equivalence class between LLR following model (1) and (2) is exhibited with mutual information as equivalence relation. Exit Charts are revisited with model (2).

A. Definition

The mutual information between L_y and L_z is defined as

$$I(L_y, L_z) = \frac{1}{log(2)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{L_y,L_z}(\ell_y, \ell_z) \times \log\left(\frac{p_{L_y,L_z}(\ell_y,\ell_z)}{p_{L_y}(\ell_y)p_{L_z}(\ell_z)}\right) d\ell_y d\ell_z \tag{3}$$

Based on the three properties above (Symmetry, Generalized Consistency and Range), $I(L_y, L_z)$ can be written as a function of the mutual information between X and the LLR as:

$$I(L_y, L_z) = I(L_y, X) + I(L_z, X) - I(L_y + L_z, X) \quad (4)$$

where $I(L, X) = 1 - \int_{-\infty}^{+\infty} p_L(\ell|X = 1) \log_2(1 + e^{-\alpha\ell}) d\ell = 1 - \int_{-\infty}^{+\infty} p_{\alpha L}(\ell|X = 1) \log_2(1 + e^{-\ell}) d\ell$. The proof is given in section A of the appendix. Some interesting properties can be derived in the special case where $I(L, X)$ is a function of a single parameter. In that case, $I(L_y, L_z)$ is a function of two variables as

$$I(L_y, L_z) = J(a_{L_y}) + J(a_{L_z}) - J(u(a_{L_y}, a_{L_z})) \quad (5)$$

where $a_{L_y}, a_{L_y} \in \mathbb{R}^+$ and $u : \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+$. Models (1) and (2) fall under this case, this is made precise below.

Example 1 [2], [19], [23]: L is the Gaussian variable in model (1) then

$$I(L, X) = J(\sigma) \quad (6)$$

with

$$J(\sigma) = 1 - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\ell - \frac{\sigma^2}{2}\right)^2}{2\sigma^2}} \log_2(1 + e^{-\ell}) d\ell \quad (7)$$

and J is a monotonically increasing function of σ [24]. The mutual information $I(L_y, L_z)$ is given by (5) with $a_{L_y} = \sigma_{L_y}$, $a_{L_z} = \sigma_{L_z}$ and $u(a_{L_y}, a_{L_z}) = \sqrt{a_{L_y}^2 + a_{L_z}^2}$.

Example 2 [17]: L is the Gaussian variable in model (2) then

$$I(L, X) = J(\alpha\sigma) \quad (8)$$

where J is the function defined in (7). We can observe that even if L is described by two parameters, $I(L, X)$ is again a monotonically increasing function of a single parameter $\alpha\sigma$. The mutual information $I(L_y, L_z)$ is given by (5) with $a_{L_y} = \alpha_{L_y}\sigma_{L_y}$, $a_{L_z} = \alpha_{L_z}\sigma_{L_z}$ and $u(a_{L_y}, a_{L_z}) = \sqrt{a_{L_y}^2 + a_{L_z}^2}$. The proof is given in Appendix A. As a consequence, it is assumed here that J is a monotonically increasing function and is thus reversible and that $u(a_{L_y}, a_{L_z}) \geq \max(a_{L_y}, a_{L_z})$. These two properties hold for example 1 and 2 above. The properties of $I(L_y, L_z)$ are listed below.

- $\max(J(a_{L_y}), J(a_{L_z})) \leq J(\sqrt{a_{L_y}^2 + a_{L_z}^2}) \leq J(a_{L_y}) + J(a_{L_z})$
- $\lim_{a_{L_y} \to 0} I(L_y, L_z) = \lim_{a_{L_y} \to 0} I(L_y, X)$
- $\lim_{a_{L_y} \to \infty} I(L_y, L_z) = I(L_z, X)$ for a given a_{L_z}
- $I(L_y, L_z) \leq min(I(L_y, X), I(L_z, X))$
- If $I(L_y, L_z) = m$ then $I(L_y, X) \geq m$ and $I(L_z, X) \geq m$.

The last property proves that $I(L_y, L_z)$ is a lower bound of the mutual information between the extrinsic and the message X. Thus if, at the end of the iterative process, $I(L_y, L_z)$ is almost equal to 1 so are $I(L_y, X)$ and $I(L_z, X)$.

From this section we can conclude that, the mutual information between extrinsics $I(L_y, L_z)$ is therefore a performance indicator for the whole system and can be used as a stopping criterion or as an error-free sequence indicator at the receiver side, provided that $I(L_y, L_z)$ can efficiently be computed online. This is addressed in section IV-A.

B. On equivalent LLR classes

Let X denote a random variable with equiprobable values in $\{-1; +1\}$. Let $G(X)$ denote the set of random variables following the model in (2) with $\alpha \in \mathbb{R}^+$ and $\sigma \in \mathbb{R}^+$. Let $G_{\alpha_0}(X)$ denote a subset of $G(X)$ such that $\alpha = \alpha_0$ is a given number in $\mathbb{R}^+$. Thus, $G(X) = \bigcup_{\alpha_0 \in \mathbb{R}^+} G_{\alpha_0}(X)$. The LLR considered in (2) spans $G(X)$ whereas the LLR considered in (1) and in density evolution or in EXIT charts spans $G_1(X)$. We first characterize the correspondence between $G(X)$ and $G_1(X)$ based on the mutual information $I(L, X)$.

Result 1: Let $L \in G(X)$. Let $L_\alpha = \alpha L$. Then $I(L, X) = I(L_\alpha, X)$ and $L_\alpha \in G_1(X)$.

Proof: $L_\alpha = \alpha^2 \frac{\sigma^2}{2} X + \alpha\sigma n$ and belongs by construction to $G_1(X)$. From (8), $I(L, X) = J(\alpha\sigma)$ and from (6), $I(L_\alpha, X) = J(\alpha\sigma)$. ■

With the notations in result 1, αL is the unique element in $G_1(X)$ with mutual information with X equal to $I(L, X)$. We prove below the stability of $G_1(X)$ under addition.

Result 2: If $L_a \in G_1(X)$ and $L_b \in G_1(X)$ then $L_a + L_b \in G_1(X)$.

Proof: $L_a = \frac{\sigma_a^2}{2} X + \sigma_a n_a$ and $L_b = \frac{\sigma_b^2}{2} X + \sigma_b n_b$ where n_a and n_b are two independent Gaussian variables with mean 0 and unit variance. $L_a + L_b = \frac{\sigma_a^2 + \sigma_b^2}{2} X + \sqrt{\sigma_a^2 + \sigma_b^2} n$ where n is a Gaussian variable with mean 0 and unit variance. ■

As a consequence, if the extrinsic LLRs of the constituent components are in $G_1(X)$, the a posteriori log-ratio is also in $G_1(X)$. From result 1, it is clear that the scaling has no impact on the mutual information of the actual LLR vector with the message. However it was observed in [14], [15], [16] and in many other references that, if a proper scaling is not used, sub-optimal LLRs may be obtained in future iterations and propagated resulting in worse performance. This point is illustrated below. The system under consideration is a serially-concatenated turbo-code (SCTC) with a $(5, 7)_8$ outer code. The inner code is a convolutional code with generator $\frac{1}{1+D}$. This SCTC will be used through the paper as an illustrative example. The two decoders are considered separately. Let L_{prior} denote the LLR at the input of a decoder with $L_{prior} \in G_\alpha(X)$. The mutual information $I(L_{prior}, X)$ is kept fixed while α is increased from 0.4 to 2.5. The mutual information is measured at the output of each decoder and plotted as a function of α in Fig. 2. We observe for the inner code that, independently of the fixed value of $I(L_{prior}, X)$, the mutual information at the output reaches a maximum when $L_{prior} \in G_1(X)$ ($\alpha = 1$). For the outer code, a maximum is also observed for low value of $I(L_{prior}, X)$ at $\alpha \approx 1$. We can conclude that EXIT charts give a prediction of the highest mutual information that can be obtained at the output of a decoder for a given $I(L_{prior}, X)$ and that this value can be reached provided that the LLR are properly scaled in order to be in $G_1(X)$. Note that the inner decoder receives as an input the LLR computed from the received data which are in $G_1(X)$ whereas L_{prior} is the sole input of the outer decoder which explains the difference observed in Fig. 2 between the inner and outer decoder.

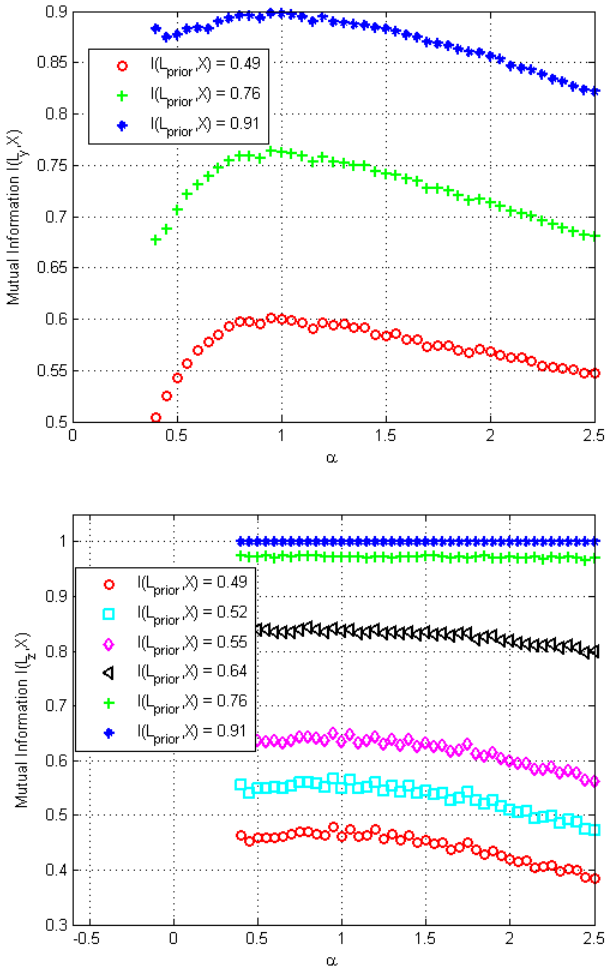


Figure 2. SCTC - Inner decoder (up) - Outer decoder (down) $EbN0 = 2dB$

C. Offline estimation : single curve exit charts

This section is devoted to the offline analysis of an iterative decoder. The performance of the association of two component decoders is usually evaluated by tracking the mutual information $I(L, X)$ at the output of each decoder considered separately and for different values of the mutual information at the input. For tracking the evolution of $I(L_y, L_z)$, both components decoders must be considered jointly. Following [19], a single curve EXIT chart (SC-EXIT) is considered here. The principle is described below for turbo and LDPC codes.

1) SC-EXIT for turbo-codes: The principle of SC-EXIT is depicted in Fig. 3. The input (a priori) is modeled with a Gaussian random variable with known mutual information $I(L, X) = J(\sigma)$ with X and is chosen in $G_1(X)$. The process is repeated for different a priori corresponding to different values of σ in the range of interest. The a priori enters decoder 1. The output L_y of decoder 1 is scaled with α_y such that $I(\alpha_y L_y, X) = I(L_y, X)$ and $\alpha_y L_y \in G_1(X)$. Then $\alpha_y L_y$ is given as an input to decoder 2. The output of decoder 2 is L_z, the corresponding scaled LLR is $\alpha_z L_z \in G_1(X)$ with parameter σ_z (σ in (1)). The mutual information $I(L_y, L_z)$ between the extrinsics is evaluated using (4) through an histogram method. The result is denoted $I_M^{(hist)}$ in the following. The computation of α is trivial since X is known. The SC-EXIT gives the evolution of $I(L_y, L_z)$ as a function of σ_g^2 and also the evolution of σ_z^2 as a function of σ_g^2. The system is able to recover the message X if $I(L_y, L_z)$ is an increasing function and if the maximum value is reached or equivalently if $\sigma_z^2 > \sigma_g^2$ [19]. If this condition is not met an oscillatory behaviour may be observed or convergence towards a local minimum may occur. In standard EXIT Charts, the curves of the individual decoders intersect. The SC-EXIT proposed here is very similar to the one in [19] except that $I(L_y, L_z)$ is not considered in [19] (tracking of the variance only) and that in [19] the non-consistency of the LLR is observed without further analysis. The translation of the SC-EXIT into an efficient stopping criterion is straightforward. This will be made explicit in the examples of section V.

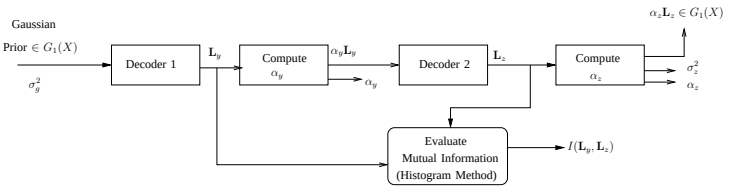


Figure 3. SC-EXIT Chart scheme for turbo-codes

2) SC-EXIT for LDPC codes: An LDPC code with m parity-check equations and length-K codewords is considered here. A soft decision algorithm such as sum-product [25] or min-sum [26], [27] is assumed at the receiver. Both algorithms are based on the exchange of extrinsics across the iterations. Tracking the evolution of $I(L_y, L_z)$ is thus also meaningful here for performance evaluation. The decoding rule is briefly recalled and connection with notations L_y and L_z is also given. Denote by $R = (R_1, R_2, ..., R_K)$ the LLR of the received signal. Denote by E_{ji} the extrinsic LLR from check node j to bit node i and M_{ji} the extrinsic LLR from bit node i to check node j. M_{ji} and E_{ji} are not defined if parity-check node j and variable i are not connected in the Tanner graph, otherwise M_{ji} and E_{ji} are updated as:

Initialization

$$M_{j,i} = R_i \tag{9}$$

Check messages

$$E_{ji} = 2tanh^{-1}\left(\prod_{i' \neq i} tanh\left(\frac{M_{ji'}}{2}\right)\right) \quad \text{Sum-Product} \tag{10}$$

$$E_{ji} = \prod_{i' \neq i} \text{sign}\left(M_{ji'}\right) \min_{i' \neq i} |M_{ji'}| \quad \text{Min-Sum} \tag{11}$$

Variable messages

$$M_{ji} = \sum_{j' \neq j} E_{j',i} + R_i \tag{12}$$

E_{ji} can be interpreted as the check node j's opinion on the probability that $X_i = 1$. We will consider also here $E_i = \frac{1}{m_i} \sum_{j \in \mathcal{A}_i} E_{ji}$ where $\mathcal{A}_i$ is the set of check nodes connected with bit i and m_i is the cardinality of $\mathcal{A}_i$. Then, E_i can be interpreted as the opinion of the check nodes on the probability

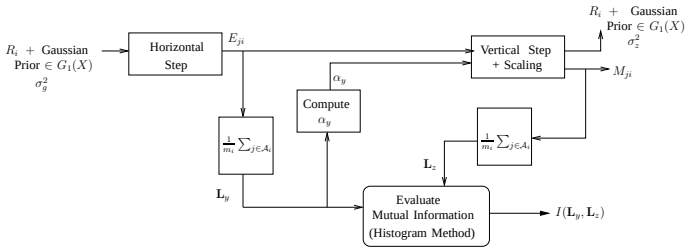


Figure 4. SC-EXIT Chart scheme for LDPC

that $X_i = 1$. In the same way, M_i is defined as $M_i = \frac{1}{m_i}\sum_{j\in\mathcal{A}_i} M_{ji}$. The SC-EXIT chart is given in Fig. 4. The input/output of the SC-EXIT is M_{ji}. To guarantee that $M_{j,i}$ remains in $G_1(X)$ it is sufficient to guarantee that $\sum_{j'\neq j} E_{j'i} \in G_1(X)$ since $R_i \in G_1(X)$ (perfect knowledge of channel characteristics or mismatch correction) and $G_1(X)$ is stable under addition. This is the role of α_y. Let L_y (resp L_z) denote the random variable associated to E_i (resp. to M_i) then L_y and L_z are noisy variables on the message X. As in the turbo-code case, the SC-EXIT Chart in Fig. 4 gives the evolution of $I(L_y, L_z)$ as a function of σ_g^2 and also the evolution of σ_z^2 as a function of σ_g^2.

IV. Efficient computation for finite-length sequences

A. Online computation of the mutual information

Obviously, the convergence behavior of the algorithm involves global quantities (on the whole sequence). Therefore, we are interested in evaluating the mutual information averaged over the whole sequence. As a first step, we consider Q frames each of length B such that $K = BQ$. The average mutual information between two extrinsics is defined as

$$I_M := \frac{1}{Q}\sum_{k=1}^{Q} I(L_{Y_k}, L_{Z_k}) \tag{13}$$

The average extrinsic information is defined as

$$I_E := \frac{1}{Q}\sum_{k=1}^{Q} I(L_k, X_k) \tag{14}$$

We demonstrate below that the average mutual information can be computed based on quantities available at the receiver, namely the extrinsics and the a posteriori probabilities. First recall two key results from [21], [28]. These results (Lemma 1 and Theorem 1) are given in [28] for binary sequences and extended to non-binary iterative decoding in [21]. They are recalled below for the binary case which corresponds to our setting.

Lemma 1: Let $\ell_{y,k}$ be the log ratio of an extrinsic probability at the output of a constituent decoder, let $\ell_{z,k}$ denote the log ratio of an a priori probability at the input of the same decoder and let $\mathbf{y}$ denote the observed sequence. Let $\ell_{z,[k]}$ denote the sequence ℓ_z from which index k is excluded. Then for all X_k in $\{-1;1\}$, $p(X_k = x_k|L_{Y,k} = \ell_{y,k}) = p(X_k = x_k|\mathbf{y}, \ell_{z,[k]})$. Lemma 1 proves that the extrinsics at the output of a constituent decoder contain the same amount of information on X_k than the observation and the a priori. The extrinsic LLR are computed as $\log\left(\frac{p(X_k=1|\mathbf{y},\ell_{z,[k]})}{p(X_k=0|\mathbf{y},\ell_{z,[k]})}\right)$. Based on lemma 1, we have $\log\left(\frac{p(X_k=1|\mathbf{y},\ell_{z,[k]})}{p(X_k=0|\mathbf{y},\ell_{z,[k]})}\right) = \log\left(\frac{p(X_k=1|L_{Y,k}=\ell_{y,k})}{p(X_k=0|L_{Y,k}=\ell_{y,k})}\right)$ which is equal to $\alpha L_{y,k}$ from the generalized consistency and symmetry assumptions. Hence, Lemma 1 proves that $\alpha\mathbf{L}_y$ (resp $\alpha\mathbf{L}_z$) should be propagated through the iterations instead of $\mathbf{L}_y$ or $\mathbf{L}_z$.

Theorem 1: Let $p_z(x_k) = p(X_k = x_k|\mathbf{y}, \ell_{z,[k]})$ represent an ergodic random variable for all X_k in $\{-1;1\}$, then the average mutual information between X and L_z is given by

$$I_E = 1 + \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\sum_{x_k} p_z(x_k)\log_2(p_z(x_k)) \tag{15}$$

The average extrinsic information can be obtained through time-averaging over a sequence of $K = BQ$ binary symbols. Theorem 1 and (4) lead to a practical evaluation of I_M:

$$\begin{aligned} I_M &= 1 + \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\sum_{x_k}\Bigg(p_z(x_k)\log_2(p_z(x_k)) \\ &+ p_y(x_k)\log_2(p_y(x_k)) \\ &- \frac{p_y(x_k)p_z(x_k)}{s_k}\log_2\left(\frac{p_y(x_k)p_z(x_k)}{s_k}\right)\Bigg) \end{aligned} \tag{16}$$

where $\sum_{x_k} p_y(x_k) = \sum_{x_k} p_z(x_k) = 1$ and $s_k = \sum_{x_k} p_y(x_k)p_z(x_k)$. The average mutual information can be approximated by time-averaging. The definition of the mutual information leads to another expression for an efficient computation of I_M.

$$\begin{aligned} I_M = \tfrac{1}{Q}\textstyle\sum_{k=1}^{Q}\int_{\ell_{y,k}}\int_{\ell_{z,k}} p(\ell_{y,k},\ell_{z,k}) \times \\ \log_2\left(\tfrac{p(\ell_{y,k},\ell_{z,k})}{p(\ell_{y,k})p(\ell_{z,k})}\right) d\ell_{y,k}d\ell_{z,k} \end{aligned} \tag{17}$$

$$I_M = \tfrac{1}{Q}\textstyle\sum_{k=1}^{Q} E\left[\log_2\left(\tfrac{p(\ell_{y,k},\ell_{z,k})}{p(\ell_{y,k})p(\ell_{z,k})}\right)\right] \tag{18}$$

The variable X_k can be introduced in the equation above by using the fact that

$$p(\ell_{y,k},\ell_{z,k}) = \sum_{x_k}\frac{p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})}{p(x_k)}p(\ell_{y,k})p(\ell_{z,k})$$

Then the average mutual information reads

$$I_M = \frac{1}{Q}\sum_{k=1}^{Q} E\left[\sum_{x_k}\log_2\left(2\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right)\right] \tag{19}$$

By assuming that the random variables involved in I_M are ergodic, we have

$$\begin{aligned} I_M &= \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\log_2\left(2\sum_{x_k} p(x_k|\ell_{y,k})p(x_k|\ell_{z,k})\right) \\ &= \lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\log_2(2s_k) \end{aligned} \tag{20}$$

By using (16) and (20), we obtain

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{x_k=0}^{1} \Big(p_z(x_k) \log_2(p_z(x_k)) + p_y(x_k) \log_2(p_y(x_k)) - \frac{p_y(x_k)p_z(x_k)}{s_k} \log_2\Big(p_y(x_k)p_z(x_k)\Big)\Big) = 0 \qquad (21)$$

At this step, two different methods can be used for an online evaluation of I_M. The first method is based on equation (16) evaluated over a sequence of finite length K. This method, called $I_M^{(a)}$, requires the knowledge of the extrinsic probabilities. The computational complexity in terms of number of operations grows linearly with K. The second method is based on equation (20) evaluated over a sequence of finite length K. This method, called $I_M^{(b)}$, has the same characteristics than $I_M^{(a)}$ with the additional advantage of a lower complexity.

The accuracy of these estimators is addressed here through an example. The system under consideration is the serially-concatenated turbo-code (SCTC) of section III-B. The block-length K is a parameter which is increased from 500 to 8000. In this example, $EbN0 = 1dB$ or $EbN0 = 2dB$ (bounds of the waterfall region). We use as a reference the exact expression of I_M in (4) evaluated through a standard histogram method as in section III-C1 and denoted $I_M^{(hist)}$. The comparison involves $I_M^{(a)}$ and $I_M^{(b)}$. The Mean Square Error (MSE) is plotted in Fig. 5. It can be observed that (20) is a better approximation of I_M than (16) which is impaired by the slow convergence of (21). If a precision of 10^{-2} is required on I_M, (20) is an accurate approximation provided that $K \geq 1500$. This precision can not be reached with (16) as long as $K \leq 10000$. For all these reasons, I_M will be approximated at the receiver side by $I_M^{(b)}$ with

$$I_M \approx I_M^{(b)} = 1 + \frac{1}{K} \sum_{k=1}^{K} \log_2(s_k) \qquad (22)$$

The standard average mutual information I_E could also be computed online through approximation (15) giving $I_E^{(a)}$. This approximation is compared with the value given by a histogram method ($I_E^{(hist)}$). The MSE is also plotted in Fig. 5. We can again observe the slower convergence of (15) with the block-length compared to (20).

We proved in this section that the mutual information $I(L_y, L_z)$ can be efficiently estimated online (at the receiver) through (22) without requiring estimation of X_i (hard decisions). The computational complexity of this approximation is linear with K. The mutual information $I(L, X)$ can also be estimated online in a similar manner but with lower accuracy. It is clear that the stopping rule and error detection device should be based on $I(L_y, L_z)$ rather than on $I(L, X)$. Note that the results in this section hold for both Gaussian and non-Gaussian distributed LLRs.

B. Online estimation of α and σ

The scaling factor and its estimation was considered in [14], [15], [16] and in many other references. In most of them, the optimal scaling factor is computed offline and for a given

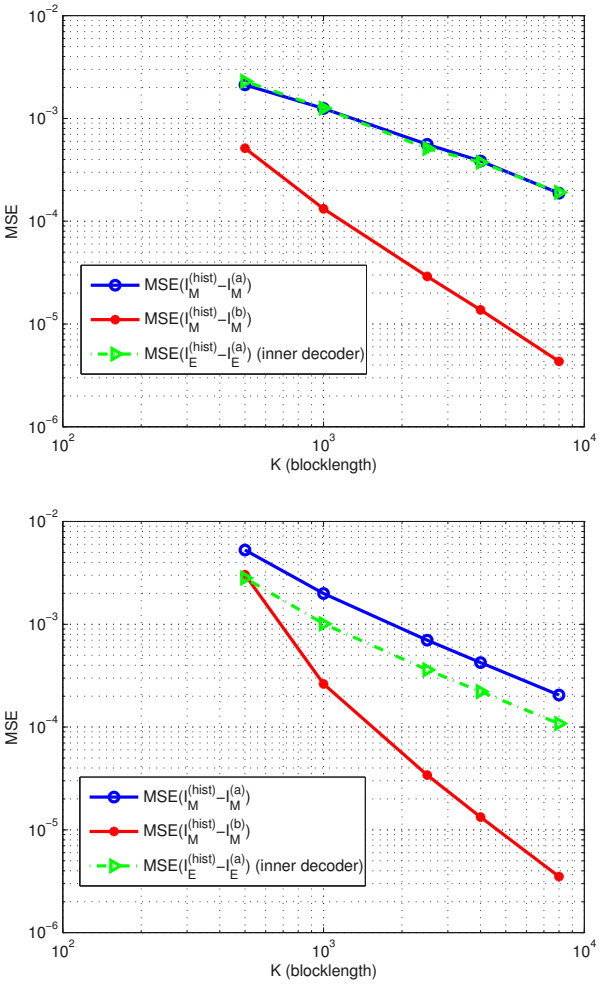


Figure 5. MSE of the difference between $I_M^{(hist)}$ and $I_M^{(a)}$ or $I_M^{(b)}$. MSE of the difference between $I_E^{(hist)}$ and $I_E^{(a)}$ at the output of the inner decoder with $EbN0 = 1dB$ (up) or $EbN0 = 2dB$ (down).

SNR value. These methods are not applicable if the channel conditions are not known with sufficient accuracy at the receiver. Recently, an online estimation method was proposed in [16] for BICM requiring an online numerical search. This method is general and does not require the LLR to be Gaussian distributed or symmetric. In [20], online LLR scaling is also derived for SNR mismatch compensation. The effectiveness of these two methods depend on the reliability of the decoder decisions which may be inaccurate during the first iteration. In the following result, an online estimation method for α and σ is provided. In contrast with most references, closed-form expressions are derived that do not require accurate decoder decisions. Our method is designed to be efficient in the whole range of SNR and at any stage of the iterative process.

Result 3: Let $L \in G(X)$ with parameters α and σ. Let $I = I(L, X)$ then

$$\sigma^2 = \frac{V(L)}{1+0.25\Big(J^{-1}(I)\Big)^2} \qquad (23)$$

$$\alpha = \frac{J^{-1}(I)}{\sigma} \qquad (24)$$

where J^{-1} is the inverse function of J in (7) and $V(L)$ stands for the variance of L.

Proof: From (2), and considering that X and n are independent random variables, $V(L) = \alpha^2 \frac{\sigma^4}{4} + \sigma^2$. Since $I = J(\alpha\sigma)$ and J is reversible, $\alpha\sigma = J^{-1}(I)$ leading to $V(L) = \left(\frac{(J^{-1}(I))^2}{4} + 1\right)\sigma^2$ which completes the proof. ■

The method in result 3 can be extended to non-Gaussian situation provided that $I(L, X)$ is a non-decreasing function of a single variable and provided that L is given by (2) in which n is not necessarily Gaussian. In the Gaussian case, $J^{-1}(I)$ can be approximated as in [29].

V. Numerical experiments

A. Example 1 : Turbo-Code

As an example, the SCTC described in section III-B is again considered here. It is explained how the mutual information can help for choosing efficiently a stopping criterion. Then, the online estimation of the scaling factor is applied to the SCTC with comparison to alternative methods from the literature. Finally, both scaling factor estimation and stopping rule are included at the decoder, the performance of the decoder is evaluated in terms of BER and number of iterations.

1) SC-EXIT chart and early stopping criterion: The SC-EXIT, for this particular system, are given in Fig. 6 when $EbN0 = 1dB$ or $EbN0 = 2dB$. We observe the following. When $\sigma_g^2 > 5$, the convergence condition ($\sigma_z^2 > \sigma_g^2$) towards the correct decision is fulfilled even when $EbN0 = 1dB$ whereas an oscillatory behavior may be observed when $\sigma_g^2 \leq 5$ preventing the convergence of the iterative process. But at the opposite, if this threshold ($\sigma_z^2 > 5$) is reached the iterative process is likely to converge in few iterations. We take advantage of this fact and propose an accurate stopping criterion as follows. The iterative process for this SCTC should be stopped when either (SC1) or (SC2) below are met:

$$(SC1) \quad I(L_y, L_z) > 1 - \epsilon \quad \text{or} \quad i > i_{max} \tag{25}$$

$$(SC2) \quad i = i_{early} \quad \text{and} \quad \sigma_z < \sqrt{5} \tag{26}$$

where ϵ is a threshold to be fixed, i_{max} is the maximum number of iterations and i_{early} is the number of iterations for deciding of an early stop. The first condition (SC1) is designed to detect error free sequences whereas (SC2) is intended to stop the process at an early stage when the algorithm is unlikely to converge to the global solution. The equivalence between different forms of the stopping rule is given below.

Result 4: Let $L \in G_1(X)$ with parameter σ, the following inequalities are equivalent :

$$\sigma > \sigma_0 \tag{27}$$

$$I(L, X) > J(\sigma_0) \tag{28}$$

$$E(|L|) > M(\sigma_0) \text{ with } M(\sigma_0) = \sigma_0\sqrt{2\pi}e^{\frac{-\sigma_0^2}{8}} + \frac{\sigma_0^2}{2} erf\left(\sqrt{\frac{\sigma_0^2}{8}}\right) \tag{29}$$

$$E(L^2) > \frac{\sigma_0^4}{4} + \sigma_0^2 \tag{30}$$

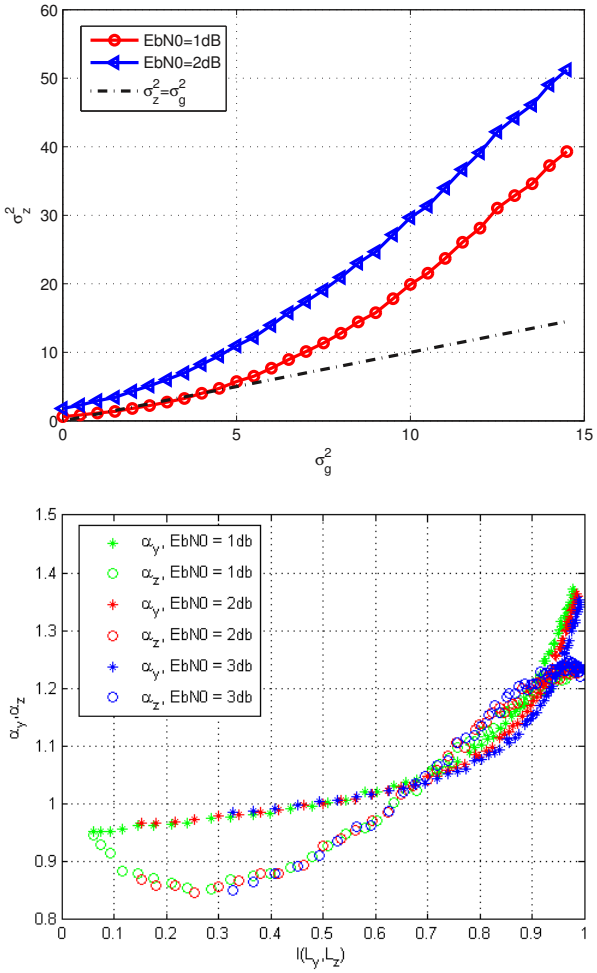


Figure 6. SC-EXIT Chart for the SCTC, up : σ_z^2 vs σ_g^2, down: α_y and α_z vs $I(L_y, L_z)$.

Let $L_y, L_z \in G_1(X)$. If $I(L_y, L_z) > 1 - \epsilon$ then $I(L_y, X) > 1 - \epsilon$ and $I(L_z, X) > 1 - \epsilon$.

Proof: Eq. (28) is equivalent to (27) because J is a strictly increasing function. When $L \in G_1(X)$, simple computation leads to $E(|L|) = M(\sigma)$ where M is a strictly increasing function which proves the equivalence between (29) and (27). The equivalence between (30) and (27) is obvious. The latter statement comes from the properties of $I(L_y, L_z)$. ■

Result 4 proves the equivalence between various stopping criterion commonly used in practice. Stopping rules (27-30) are based on the reliability of the LLR at the output of a given decoder. Alternative methods such as Cross-entropy [7] measure the similarity of the distributions at the output of the two decoders. Once the system has converged to the correct solution, the magnitude of the LLRs has reached a high level (reliability) and the two decoders agree on the decisions (similarity). It is proved below that I_M the estimator of $I(L_y, L_z)$ reaches its maximum if and only if the two conditions are met.

Result 5: $I(L_y, L_z) \approx I_M = 1$ if and only if $\text{sign}(L_{y,i}) = \text{sign}(L_{z,i})$ and $|L_{y,i}| = |L_{z,i}| = +\infty \ \forall i \in \{1, 2..., K\}$.

Proof: $I_M = 1 + \frac{1}{K}\sum_{i=1}^{K} \log_2\left(\frac{1+e^{L_{y,i}+L_{z,i}}}{(1+e^{L_{y,i}})(1+e^{L_{z,i}})}\right)$ and $I_M = 1$ is equivalent to $\frac{1+e^{L_{y,i}+L_{z,i}}}{(1+e^{L_{y,i}})(1+e^{L_{z,i}})} = 1\ \forall i \in \{1,2...,K\}$. This condition is met if and only if $\text{sign}(L_y,i) = \text{sign}(L_z,i)$ and $|L_y| = |L_z| = +\infty\ \forall i \in \{1,2...,K\}$. ■

As an illustration, the shape of I_M and of cross-entropy (CE) are compared in Fig. 7 for a single bit with probability measures p_y anf p_z. We can seen that $1-CE$ reaches its maximum when $p_y(x_k) = p_z(x_k)$ (similarity) whereas I_M reaches its maximum when $p_y(x_k) = p_z(x_k) = 0$ or $p_y(x_k) = p_z(x_k) = 1$ (reliability and similarity).

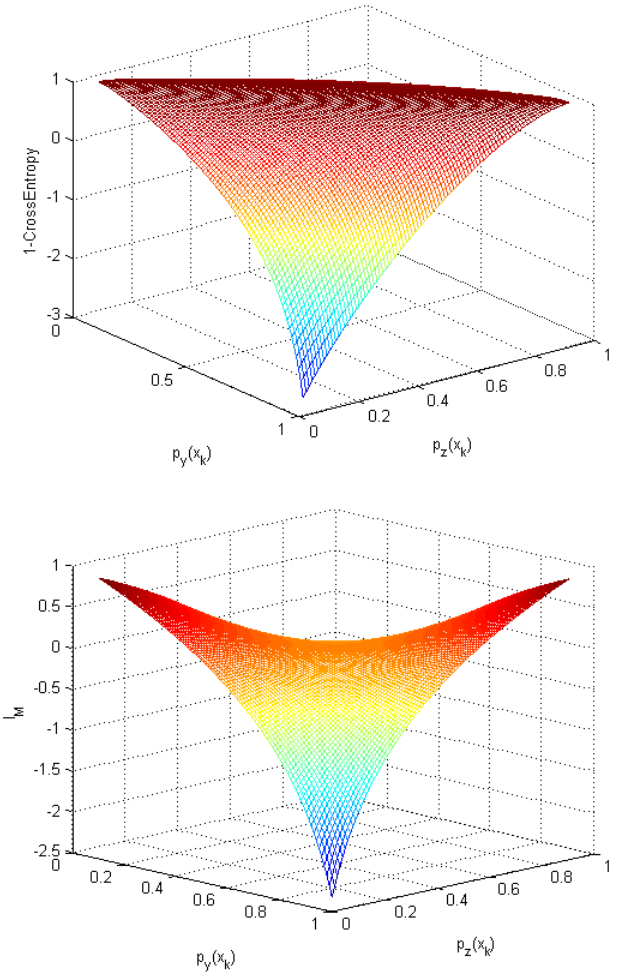


Figure 7. Comparison between $1 - CE$ (up) and I_M (down).

From the SC-EXIT, the typical values of α_y and α_z can also be obtained. They are plotted in Fig. 6 as a function of $I(L_y, L_z)$ and for different values of the SNR. We observe that the value of α varies with the iteration number and with the SNR but this information is captured in $I(L_y, L_z)$. As a consequence, the scaling factor α is a function of a single parameter $I(L_y, L_z)$. We conclude that an alternative method for estimating the value of α consists in evaluating online the mutual information $I(L_y, L_z)$ with (22) and using Fig. 6 to obtain the corresponding α_y and α_z. The coefficients of the polynomials that fits each of the two curves $\alpha_y = f(I(L_y, L_z))$ and $\alpha_z = g(I(L_y, L_z))$ in the least square sense are given in Appendix B.

2) Scaling factor estimation: Two methods for an online estimation of the scaling factor α are proposed in this paper. The efficiency of these methods is measured here for the SCTC. Comparison is also provided with two other methods from the literature:

- Generalized Mutual Information (GMI) [16]. The optimal scaling factor $\hat{s}$ reads:

$$\hat{\alpha} = \arg\min_{\alpha}\{1 - \frac{1}{K}\sum_{i=1} \log_2(1 + \exp(-\text{sign}(\hat{X}_i L_i \alpha)))\}$$

  where $\hat{X}_i$ is the hard decision of the decoder that provides an estimate for X_i. This online method maximizes the generalized mutual information and is proved to be equivalent to offline explicit method for deriving the scaling factor. The method does not require the Gaussian assumption to hold. The scaling factor is obtained from online numerical search (iterative method). The accuracy depends on decoder decision and may be inaccurate at low SNR and at the beginning of the iterative process.
- LLR absolute value [30]. It is assumed here that L is Gaussian distributed as in (2). The scaling factor is obtained by measuring the mean $E(|L|)$ and variance $V(|L|)$ of the absolute value of the LLR and by solving the system of equations below :

$$V(|L|) = \frac{\alpha^2\sigma^4}{4} + \sigma^2 \quad (31)$$

$$\frac{V(|L|)}{E^2[|L|]} = \frac{1+\frac{\alpha^2\sigma^2}{4}}{\left(\sqrt{\frac{2}{\pi}}e^{-\frac{\alpha^2\sigma^2}{8}} + \frac{\alpha\sigma}{2}erf(\frac{\alpha\sigma}{2\sqrt{2}})\right)^2} = h(\alpha\sigma) \quad (32)$$

  A second order polynomial is found in [30] to approximate $h^{-1}\left(\frac{V(|L|)}{E^2[|L|]}\right)$ in the range $0 - 6dB$ avoiding numerical integration.
- Mutual information (MI). This is the method in result 3 where I reads (Eq. (15)):

$$I = I(L,X) \approx 1 + \frac{1}{\log(2)K}\sum_{i=1}^{K}\left(\frac{L_i e^{L_i}}{1+e^{L_i}} - \log(1+e^{L_i})\right)$$

  This method holds as long as J is a strictly increasing function of the variable $\alpha\sigma$. In this paper, J is obtained by assuming Gaussian distributed LLR and $J^{-1}(I)$ is computed with the approximation in the appendix of reference [29].
- SC-EXIT. The scaling factor is estimated with (39) for the inner code and (40) for the outer code. From Eq. (22), the mutual information $I(L_y, L_z)$ reads:

$$I(L_y, L_z) \approx 1 + \frac{1}{K}\sum_{i=1}^{K}\log_2\left(\frac{1+e^{L_{y,i}+L_{z,i}}}{(1+e^{L_{y,i}})(1+e^{L_{z,i}})}\right)$$

The accuracy of the proposed methods is quantified with the normalized mean error as $E(\frac{|\hat{\alpha}-\alpha|}{\alpha})$ where $\hat{\alpha}$ is the estimate provided by one of the methods under consideration and α is the desired scaling factor. The latter is obtained assuming true transmitted bits are available. The MSE is plotted in Fig. 8 as a function of the mutual information in order to

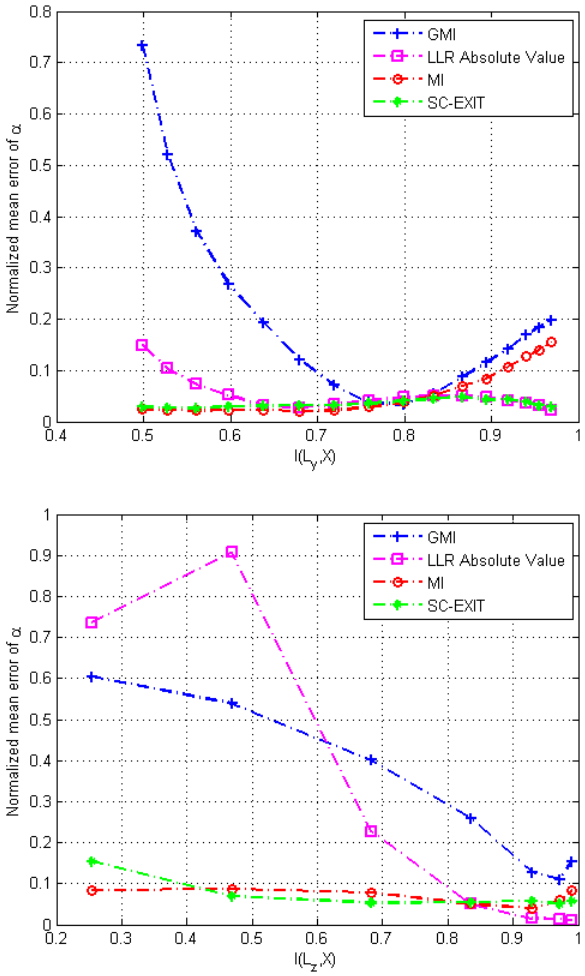


Figure 8. Scaling factor estimation - Output of inner code (up) - Output of outer code (down).

evaluate the efficiency of the method for various SNR and at different stages of the iterative process. As expected, GMI is not sufficiently accurate for low values of the mutual information and is dismissed here since accurate estimation of α is mandatory in early iterations (see Fig. 2). The method in [30] is very sensitive to violation of the Gaussian assumption. This may occur in early iterations especially at the outer code that does not admit the channel observations as an input. At the opposite, we observe that the two methods proposed in this paper are valid over the whole range of SNR and are robust, to a certain extent, to non-Gaussian data. It can be noticed that the Gaussian assumption is not required in the derivation of (15) and (22).

3) BER curves: The optimal parameters deduced from the SC-EXIT are now included at the decoder side. Precisely, we now consider as stopping criterion (SC1) in (25) combined with the early stopping criterion (SC2) in (26) with $i_{early} = 15$. We proved in the previous section that stopping rules can take several equivalent expressions. Two configurations are considered here: *(a)* (SC1) and (SC2) involve $I(L_y, L_z)$ as stopping indicator and α_y and α_z are estimated with the method called SC-EXIT in the previous section, *(b)* (SC1) and (SC2) involve σ as stopping indicator and α_y and α_z are estimated with the method in (24). The thresholds for (SC2) are based on the EXIT curve ($\sigma \leq \sqrt{5}$, $I(L_y, L_z) \leq 0.4$) whereas the thresholds in (SC1) are chosen in order to obtain the same performance in terms of BER for the two methods ($\epsilon = 10^{-2}$ and $\sqrt{\sigma_y^2 + \sigma_z^2} > 20$). The performance is evaluated in terms of BER and WER and is plotted in Fig. 9 with label $SCTC_{optim}$. Comparison is given with the standard decoder ($SCTC_{standard}$) with unscaled LLR and a fixed number of iterations ($n = 15$). Two different codeword lengths are considered: $K = 1000$ and $K = 8000$. In both cases, we observe a slight improvement in the performance with $SCTC_{optim}$. This is due to the propagation of scaled LLR and also to a better repartition of the iterations thanks to the stopping criterion. The average number of iterations is reported in Fig. 9 confirming the validity of (SC1) as an accurate stopping rule based on a performance criterion (error free). At low $EbN0$, stopping rule (SC2) maintains the average number of iterations at a reasonable level even if the maximum number of iterations allowed is large ($i_{max} = 50$ here) with a better efficiency when (SC2) is based on $I(L_y, L_z)$.

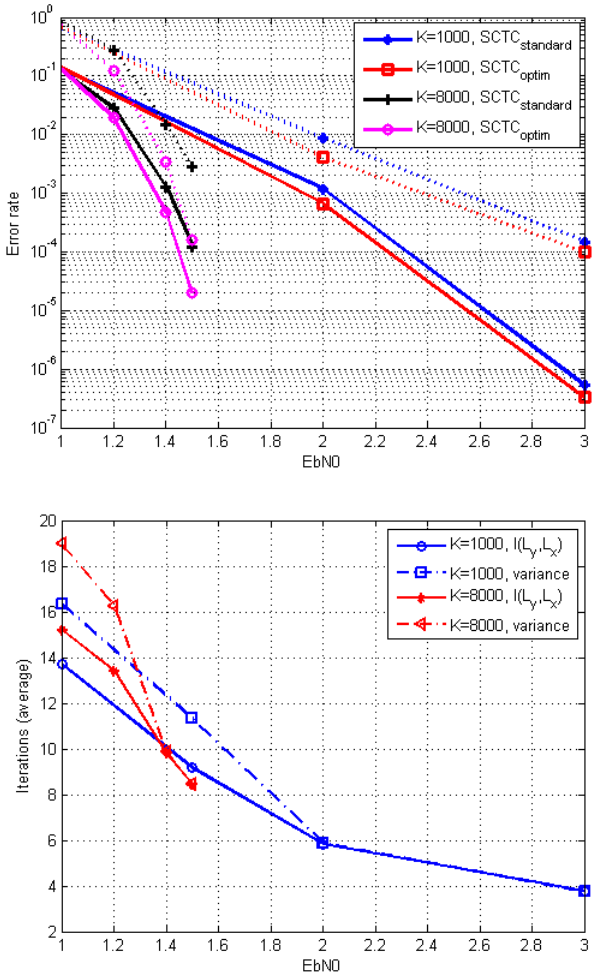


Figure 9. SCTC with $K = 1000$ or $K = 8000$ - Up: BER (solid line), WER (dotted line)- Down: Average number of iterations.

B. Example 2 : LDPC codes

In this section, a (3-6) regular LDPC code and a (4,8) regular LDPC are considered. The SC-Exit charts are given in the section below.

1) SC-Exit chart: The SC-EXIT Chart in Fig. 4 is applied to regular LDPC codes. The evolution of $I(L_y, L_z)$ is plotted in Fig. 10 when either Sum-Product of Min-Sum is implemented. We observe that the curves obtained with Sum-Product and Min-Sum are superimposed when an optimal scaling factor is applied meaning that the two implementations yield the same performance. In several publications, Min-Sum with scaling factor is compared to Sum-Product without scaling for which a slightly worse performance is noted at low SNR. The comparison is unfair, optimal scaling with Sum-Product must correct this phenomenon. The value of the scaling factor α_y is given in Fig. 11. We can see again that α_y is a function of the single variable $I(L_y, L_z)$ since the value obtained for α_y always falls on the same curve independently of $EbN0$ and of the iteration number. From Fig. 11, we can conclude that LLR-scaling is mandatory when using Min-Sum at the decoder side. In addition, we can observe that α_y is not in the immediate vicinity of 1 even when Sum-Product is implemented at the decoder. As with Min-Sum, the LLR are overestimated in the first iterations and should be scaled to improve the performance. The performance of (3-6) and (4-8)-LDPC codes are now evaluated with different implementations at the decoder. We consider Sum-Product with either unscaled LLR (SP, $\alpha = 1$) or optimally scaled LLR (SP, α_{optim}). We also consider Min-Sum either with unscaled LLR (MS, $\alpha = 1$) or with a fixed scaling factor (SP, $\alpha = 0.8$). The value of the scale factor ($\alpha = 0.8$) is obtained from the literature [14], [15] where its optimality is assessed for regular (3-6)-LDPC codes. In this experiment, the codeword-length is $K = 1000$. The stopping criterion is $\mathbf{H}\hat{\mathbf{c}} = 0$ or $i = it_{max}$ where $\mathbf{H}$ is the parity-check matrix, $\hat{\mathbf{c}}$ is the hard-decision on the encoded bits and $it_{max} = 50$. The BER is plotted as a function of $EbN0$ in Fig. 12. We observe the following, scaling LLR always improves the BER. The gain is more important when α is far from 1 and at high $EbN0$. When properly scaled, Min-Sum and Sum-product exhibit the same performance. We can observe that $\alpha = 0.8$ is indeed optimal with (3-6)-LDPC and is close to be optimal with (4-8)-LDPC.

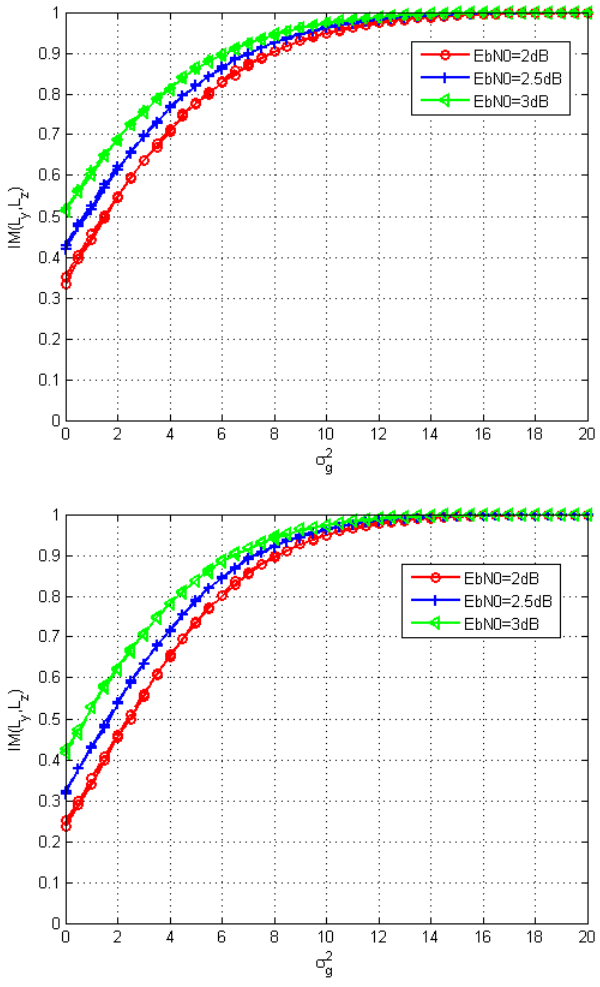


Figure 10. SC-EXIT Chart, $I(L_y, L_z)$ vs σ_g^2 for (3-6)-LDPC (up) and for (4-8)-LDPC (down)- Solid line: Sum-Product, Dotted line: Min-Sum.

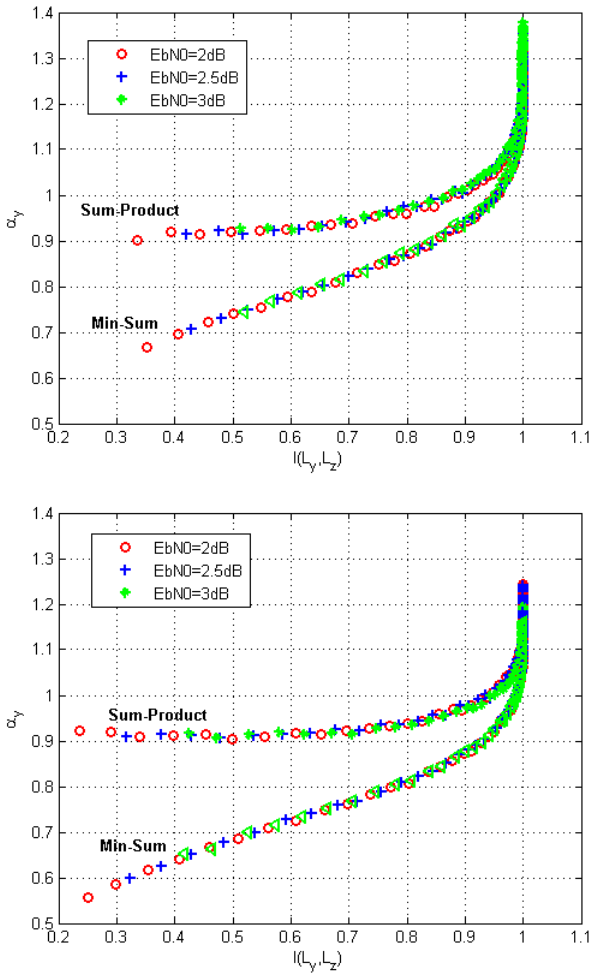


Figure 11. SC-EXIT Chart, α_y vs $I(L_y, L_z)$ for (3-6)-LDPC (up) and (4-8)-LDPC (down).

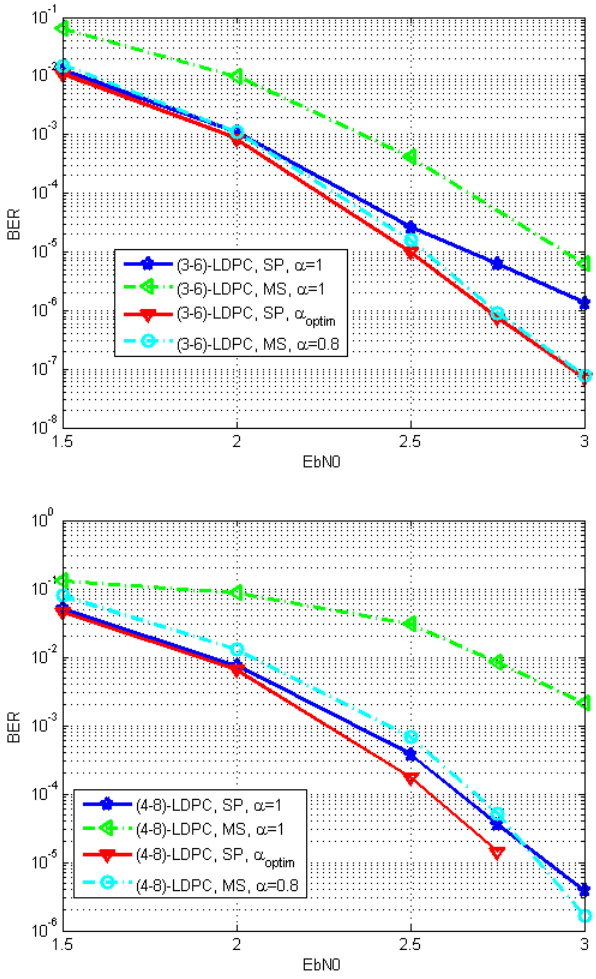


Figure 12. BER curves (3-6)-LDPC (up) and for (4-8)-LDPC (down).

VI. Conclusion

In this paper, we studied the properties of the mutual information between the extrinsics involved in an iterative decoding process. Our framework is quite general and encompasses serially or parallel concatenated turbo-codes or LDPC codes as special cases. We proved that the mutual information between extrinsics is a pertinent performance measure that can be used at the receiver side for error detection purpose. In addition, an offline evaluation, reminiscent of EXIT Charts, gives guidelines for defining efficient stopping rules. Two methods for an online efficient estimation of scaling factors are also derived. Numerical example highlighted the benefits and the generality of the proposed approach.

Acknowledgement

Thanks to Pierre Duhamel for fruitful discussions and constructive criticism.

Appendix

A. Proof of Eq. (4):

In the following, we consider that the three properties (Symmetry, Generalized Consistency and Range) hold for L_y and L_z. We will used the notation α_Y (resp. α_Z) in the Generalized Consistency property when applied to L_y (resp. L_z) and α otherwise (*ie* the result hold for both and is given for L which stands for L_y or L_z).

The expression of $p_L(\ell)$ is given by $p_L(\ell) = \frac{1}{2}\sum_x p_L(\ell|x)$. Since L_y and L_z are independent when conditioned on X, we have

$$\begin{aligned}\log\left(\frac{p_{L_y,L_z}(\ell_y,\ell_z)}{p_{L_y}(\ell_y)p_{L_z}(\ell_z)}\right) &= \log(2)\\ &+ \underbrace{\log\left(\sum_x p_{L_y}(\ell_Y|x)p_{L_z}(\ell_Z|x)\right)}_{U_{L_y,L_z}} - \underbrace{\log\left(\sum_x p_{L_y}(\ell_y|x)\right)}_{U_{L_y}}\\ &\underbrace{-\log\left(\sum_x p_{L_z}(\ell_z|x)\right)}_{U_{L_z}} \qquad (33)\end{aligned}$$

U_{L_y} and U_{L_z} have similar expression $U_L = \log\left(p_L(\ell|x)\right) + \log\left(1+e^{-\alpha x\ell}\right)$ whereas $U_{L_y,L_z} = \log\left(p_{L_y}(\ell_y|x)p_{L_z}(\ell_z|x)\right) + \log\left(1+e^{-x(\alpha_y\ell_y+\alpha_z\ell_z)}\right)$. The mutual information reads

$$\begin{aligned}I(L_y,L_z) &= 1 + \\ &\frac{1}{log_2(2)}\Big(\underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y,L_z}(\ell_y,\ell_z)U_{L_y,L_z}dl_ydl_z}_{A_{L_y,L_z}}\\ &\underbrace{-\int_{-\infty}^{+\infty} p_{L_y}(\ell_y)U_{L_y}dl_y}_{A_{L_y}} - \underbrace{\int_{-\infty}^{+\infty} p_{L_z}(\ell_z)U_{L_z}dl_z}_{A_{L_z}}\Big) \qquad (34)\end{aligned}$$

where $A_L = \frac{1}{2}\sum_x \int_{-\infty}^{+\infty} p_L(\ell|x)\log(p_L(\ell|x))d\ell + \int_{-\infty}^{+\infty} p_L(\ell|X=1)\log\left(1+e^{-\alpha\ell}\right)d\ell$ and

$$\begin{aligned}A_{L_y,L_z} &= \tfrac{1}{2}\textstyle\sum_x \int_{-\infty}^{+\infty} p_{L_y}(\ell_y|x)\log(p_{L_y}(\ell_y|x))d\ell_y\\ &+\tfrac{1}{2}\textstyle\sum_x \int_{-\infty}^{+\infty} p_{L_z}(\ell_z|x)\log(p_{L_z}(\ell_z|x))d\ell_z\\ &+\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y}(\ell_y|X=1)p_{L_z}(\ell_z|X=1)\times\\ &\log(1+e^{-(\alpha_y\ell_y+\alpha_z\ell_z)})d\ell_yd\ell_z \qquad (35)\end{aligned}$$

Then the mutual information reads

$$\begin{aligned}I(L_y,L_z) &= I(L_y,X) + I(L_z,X)\\ &+\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y}(\ell_y|X=1)p_{L_z}(\ell_z|X=1)\times\\ &\log(1+e^{-(\alpha_y\ell_y+\alpha_z\ell_z)})d\ell_yd\ell_z - 1 \qquad (36)\end{aligned}$$

We prove now that the last term in (36) is the mutual information between $L_y + L_z$ and X. We can first remark that (substitution in the integral)

$$\begin{aligned}&\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{L_y}(\ell_y|X=1)p_{L_z}(\ell_z|X=1)\times\\ &\log(1+e^{-(\alpha_y\ell_y+\alpha_z\ell_z)})d\ell_yd\ell_z =\\ &\textstyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{\alpha_yL_y}(\ell_y|X=1)p_{\alpha_zL_z}(\ell_z|X=1)\times\\ &\log(1+e^{-(\ell_y+\ell_z)})d\ell_yd\ell_z \qquad (37)\end{aligned}$$

Let $u = \ell_y + \ell_z$ and $t = \ell_z$, we have

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{\alpha_y L_y}(\ell_y|X=1)p_{\alpha_z L_z}(\ell_z|X=1)\times \log(1+e^{-(\ell_y+\ell_z)})d\ell_y d\ell_z = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} p_{\alpha_y L_y}(u-t|X=1)p_{\alpha_z L_z}(t|X=1)dt\right)\times \log(1+e^{-u})du$$

The convolution leads to

$$I(L_y,L_z) = I(L_y,X)+I(L_z,X)+\int_{-\infty}^{+\infty} p_{\alpha_y L_y+\alpha_z L_z}(\ell_y+\ell_z|1)\log_2(1+e^{-\ell_y-\ell_z})d\ell_y d\ell_z - 1 = I(L_y,X)+I(L_z,X)-I(L_y+L_z,X) \quad (38)$$

which concludes the proof.

B. Approximation of $\alpha = f(I(L_y,L_z)$ for the SCTC

The two curves in Fig. 6 can be approximated with the polynomials below:

$$\alpha_y = 5.6574x^6 - 12.3492x^5 + 10.3204x^4 - 3.7379x^3 + 0.5703x^2 + 0.0671x + 0.9461 \quad (39)$$

$$\alpha_z = -9.1055x^5 + 24.933x^4 - 26.7916x^3 + 14.8465x^2 - 3.8655x + 1.2148 \quad (40)$$

where $x = I(L_y,L_z)$.

REFERENCES

[1] S. ten Brink, "Convergence of iterative decoding," *Electronics Letters*, vol. 35, no. 13, pp. 1117–1119, 1999.

[2] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Commun.*, vol. 49, pp. 1727–1737, Oct 2001.

[3] E. Sharon, A. Ashikhmin, and S. Litsyn, "Analysis of low-density parity-check codes based on EXIT functions," *IEEE Transactions on Communications*, vol. 54, no. 8, pp. 1407–1414, 2006.

[4] S. ten Brink and G. Kramer, "Design of repeat-accumulate codes for iterative detection and decoding," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2764–2772, 2003.

[5] A. Grant, "Convergence of non-binary iterative decoding," in *GLOBECOM*, Nov. 2001, vol. 2, pp. 1058–1062.

[6] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun*, 1993, pp. 1064–1070.

[7] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.

[8] R. Shao, S. Lin, and M. Fossorier, "Two simple stopping criteria for turbo decoding," *IEEE Trans. Commun.*, vol. 47, pp. 1117–1120, 1999.

[9] Y Wu, B. Woerner, and J. Ebel, "A simple stopping criterion for iterative decoding," *IEEE Communications Letters*, vol. 4, pp. 258–260, 2000.

[10] T. Ngatched and R. Takawira, "Simple stopping criterion for iterative decoding," *Electronics Letters*, vol. 37, pp. 1350–1351, 2001.

[11] L. Kocarev, F. Lehmann, G.M. Maggio, B. Scanavino, Z. Tasev, and A. Vardy, "Nonlinear dynamics of iterative decoding systems: analysis and applications," *IEEE Trans. on Infor. Theory*, vol. 52, no. 4, pp. 1366–1384, 2006.

[12] F. Zhai and I.J. Fair, "New error detection techniques and stopping criteria for turbo decoding," in *Canadian Conference on Electrical and Computer Engineering*, 2000, vol. 1, pp. 58–62 vol.1.

[13] Fan-Min Li and An-Yeu Wu, "On the new stopping criteria of iterative turbo decoding by using decoding threshold," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5506–5516, 2007.

[14] Jinghu Chen, A. Dholakia, E. Eleftheriou, M.P.C. Fossorier, and Xiao-Yu Hu, "Reduced-Complexity Decoding of LDPC Codes," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1288–1299, Aug 2005.

[15] A. Alvarado, V. Nunez, L. Szczecinski, and E. Agrell, "Correcting suboptimal metrics in iterative decoders," in *Communications, 2009. ICC '09. IEEE International Conference on*, June 2009, pp. 1–6.

[16] Jinhong Wu, M. El-Khamy, Jungwon Lee, and Inyup Kang, "BICM performance improvement via online LLR optimization," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, April 2013, pp. 3850–3855.

[17] M. Fu, "Stochastic analysis of turbo decoding," *IEEE trans on Inform. Theory*, vol. 51, no. 1, pp. 81–100, 2005.

[18] M. Fu, "On Gaussian approximation for density evolution of low-density parity-check codes," in *IEEE International Conference on Communications, 2006. ICC '06.*, June 2006, vol. 3, pp. 1107–1112.

[19] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 891–907, 2001.

[20] M. El-Khamy, J. Wu, J. Lee, and I. Kang, "Online log-likelihood ratio scaling for robust turbo decoding," *Communications, IET*, vol. 8, no. 2, pp. 217–226, January 2014.

[21] J. Kliewer, S. Xin Ng, and L. Hanzo, "Efficient computation of EXIT functions for nonbinary iterative decoding," *IEEE Trans. on Commun.*, vol. 54, no. 12, pp. 2133–2136, December 2006.

[22] J. Hagenauer, "The EXIT chart: Introduction to extrinsic information transfer in iterative processing," in *Proceedings of Eur. Signal Process. Conf.*, Vienna, Austria, Sept 2004, pp. 1541 –1548.

[23] H. El-Gamal and Jr. Hammons, A.R., "Analyzing the turbo decoder using the gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 671–686, 2001.

[24] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[25] D. J C MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 399–431, Mar 1999.

[26] M.P.C. Fossorier, M. Mihaljevic, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Transactions on Communications*, vol. 47, no. 5, pp. 673–680, May 1999.

[27] N. Wiberg, *Codes and decoding on general graphs*, Ph.D. thesis, Link oping Univ., Link oping, Sweden, 1996.

[28] I. Land, P. Hoeher, and S. Gligorević, "Computation of symbol-wise mutual information in transmission system with log APP decoders and application to EXIT charts," in *in Proc. Int. ITG Conf. Source Channel Coding*, Erlangen, Germany, Jan. 2004, pp. 195–202.

[29] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Transactions on Communications*, vol. 52, no. 4, pp. 670–678, April 2004.

[30] T.A. Summers and S.G. Wilson, "SNR mismatch and online estimation in turbo decoding," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 421–423, Apr 1998.